

AD-A092 575

STANFORD UNIV CA DEPT OF STATISTICS

F/6 12/1

METHODS AND APPLICATIONS OF TIME SERIES ANALYSIS. PART I. REGRE--ETC(U)

JUL 80 T W ANDERSON, N D SINGPURWALLA

N00014-75-C-0442

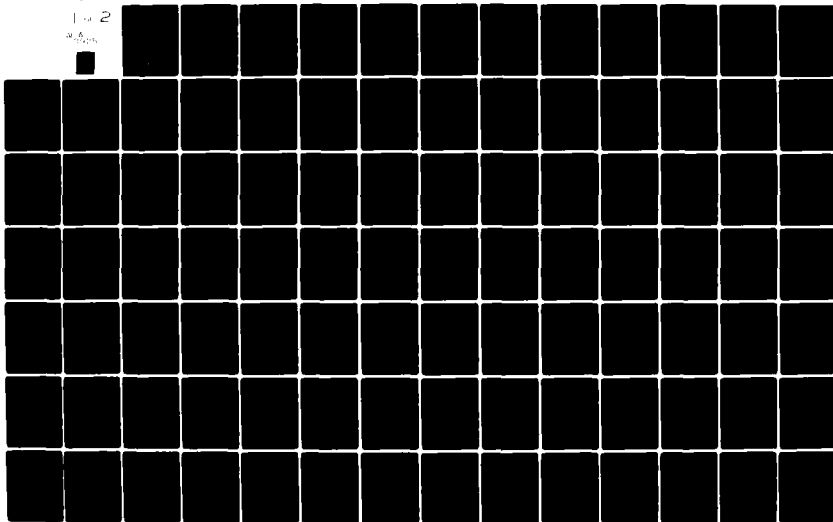
UNCLASSIFIED

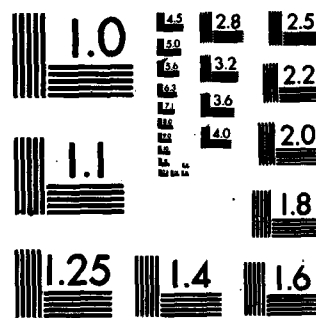
TR-42

NL

1 of 2

W. A. SINGPURWALLA





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

METHODS AND APPLICATIONS OF TIME SERIES ANALYSIS
Part I: Regression, Trends, Smoothing, and Differencing

BY
T. W. ANDERSON
and

N. D. SINGPURWALLA

TECHNICAL REPORT NO. 42 ✓
JULY 1980

PREPARED UNDER CONTRACT N00014-75-C-0442 ✓
(NR-042-034)

OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

✓ **DEPARTMENT OF STATISTICS**
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



DISTRIBUTION STATEMENT A

Approved for public release:
Distribution Unlimited

80 12 02 044

AD A092575

DDC FILE COPY

12
B.S.

201473

DTIC
ELEC
DEC 3 1980
C

METHODS AND APPLICATIONS OF TIME SERIES ANALYSIS
Part I: Regression, Trends, Smoothing, and Differencing
by

T. W. Anderson
Stanford University

and

N. D. Singpurwalla
George Washington University

TECHNICAL REPORT NO. 42

JULY 1980

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)

OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government.
Approved for public release; distribution unlimited.

Also issued as ^{NH}Serial T-426, The Institute for Management Science and
Engineering, School of Engineering and Applied Science, George Washington
University, Washington, D. C. 20052

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

© T. W. Anderson and N. D. Singpurwalla 1980

Acknowledgements

We would like to acknowledge the assistance of Dolores Conway, Marelene Egger, Gabrielle Kelley, Mohamed Habibullah and Man Yuen Wong in connection with the some of the examples discussed here. We would also like to acknowledge the enthusiasm of the students in Spring 1978 Quarter of Statistics 207 at Stanford University, who provided us with some unusual sources of data for our examples.

Accession For	
NTIS CRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

Preface

A series of technical reports on methodology in time series analysis with applications is planned. The reports will be issued sequentially, each covering an important aspect. The first is concerned with Regression, Trends, Smoothing, and Differencing. Subsequent reports will deal with forecasting, autoregressive integrated moving average models and statistical techniques based on them, serial correlation, and spectral analysis.

The purpose of this series of technical reports is to develop the most modern procedures of time series analysis and forecasting for use in engineering, the physical sciences, and the social sciences. It is expected that these techniques will be useful in scientific and managerial activities of the Department of Defense. The exposition of methodology is based on a succinct presentation of the theoretical background and is illustrated with appropriate examples from engineering, maintenance and reliability, economics, and other physical and social sciences; these will give a "real-life" flavor to the presentation.

Much of the material in these technical reports is based on The Statistical Analysis of Time Series by T.W. Anderson, the preparation of which was supported by the Office of Naval Research. A preliminary version, written by N.D. Singpurwalla, was used in the Stanford University course: Statistics 207. Introduction to Time Series Analysis. It is assumed that the reader has some background in mathematics - calculus of several variables and introductory linear algebra - and a basic statistics course.

CONTENTS

Methods and Applications of Time Series Analysis

Part I: Regression, Trends, Smoothing, and Differencing

1. Introduction	1
1.1 What Is a Time Series?	1
1.2 Examples of Time Series	1
1.3 Why Do We Study Time Series?	2
1.4 How Do We Take Measurements on a Time Series?	2
1.5 What Types of Time Series Do We Study Here?	2
1.6 A General Model	3
1.7 The Regression Function	4
1.8 Stationary Stochastic Processes	5
1.9 More General Models	7
2. The Use of Regression Analysis in Time Series Analysis	8
2.1 An Outline of the General Theory of Least Squares	8
2.2 Prediction	12
2.3 Some Illustrative Examples	13
3. Trends and Smoothing	22
3.1 Polynomial Trends	23
3.2 Smoothing	37
3.3 The Variate Difference Method	53
4. Cyclical Trends	57
4.1 Transformations and Representations	57
4.2 Statistical Estimation Procedures for Cyclical Trends	79
4.3 The Periodogram and the Spectrum	85
4.4 Tests of Hypotheses and Confidence Regions for Coefficients	90
References	98

Methods and Applications of Time Series Analysis

by

T.W. Anderson and N.D. Singpurwalla
Stanford University and George Washington University

Part I: Regression, Trends, Smoothing, and Differencing

1. Introduction

1.1 What Is a Time Series?

A time series is a sequence of observations which is ordered in time (or some other entity of interest); the measurements may be temperature, stress, Gross National Product, etc. The feature of a time series analysis that distinguishes it from other statistical analyses is the explicit recognition of the fact that the observations arrive according to some order. While in many situations the observations are assumed to be statistically independent, in time series analysis the possible dependence between the observations is given prime consideration.

1.2 Examples of Time Series

In many areas of daily interest, there are phenomena whose evolution and variation with the passing of time are of interest. Examples of these are the weather, the price of gasoline, the level of industrial output, the state of one's health, and the popularity of classical music. The measurement of any particular characteristic over time constitutes a realization of a time series. Often we are interested in measuring several characteristics over time; for example, an electrocardiogram consists of several records.

1.3 Why Do We Study Time Series?

A very pragmatic objective for studying a time series may be to predict the future based upon a knowledge of the past. Another objective may be to obtain an understanding of the mechanism producing the series, so that we may be able to control it and obtain desirable results. A less pragmatic objective might be simply to obtain a succinct description of the salient features of the series.

1.4 How Do We Take Measurements on a Time Series?

Even though many quantities, such as temperature and wind velocity, change continuously in time and can sometimes be recorded continuously in the form of a graph, very often in practice measurements are made in discrete time. Digital computers, which are used for an analysis of time series, accept data that is available at discrete points in time. In view of the above, we shall confine ourselves to time series that are recorded discretely in time at regular intervals, such as at each hour on the hour or at the end of each year.

1.5 What Types of Time Series Do We Study Here?

We shall assume that the measurements that we make in a time series are comprised of real numbers which are not limited to a finite (or countable) number of values. That is, our measurements will consist of "continuous" values. For example, the number of transistors that we manufacture per day is assumed to be so large that there is no sacrifice of reality if we consider it to be a continuous variable.

Furthermore, we shall only consider series which are stable; that is, values of these stay within certain bounds and they are changing slowly and not abruptly such as shock waves caused by an underwater explosion.

1.6 A General Model

Consider T equally spaced time points which we label as $1, 2, 3, \dots, T$; let us take observations on the time series at these time points and denote these observations by y_1, y_2, \dots, y_T . A simple-minded, yet fairly general, model for the times series can be written as

$$y_t = f(t) + u_t, \quad t = 1, 2, \dots, T.$$

That is, we say that the observed series is made up of a completely deterministic (determined) part $f(t)$, where $f(t)$ is some function of time t , and a random or a stochastic part u_t , where u_t obeys some probabilistic law. In electrical engineering $f(t)$ is often referred to as a signal, and u_t as the noise. The quantities $f(t)$ and u_t are not observable by us; they are theoretical quantities which represent our abstraction of the series y_t . For example, if the y_t denote measurements on the daily rainfall, then the $f(t)$ may represent the long-run average rainfall at day t taken over many years, and the u_t would represent the daily irregularities which describe the fluctuations from the norm. The random part u_t has the usual "frequency" interpretation that we use in statistics. That is, if in theory, we could repeat the entire situation under which the observations

y_1, y_2, \dots, y_T were obtained, (in reality time proceeds progressively in one direction), then the $f(t)$ would be the same as before, but the random terms u_t would be different. That is, the various values of u_t , at time t would be described by a frequency function. Any errors of measurement (which are usually prevalent in many physical situations) will be included in the u_t .

Our purpose in specifying the model $y_t = f(t) + u_t$ is to represent the mechanism generating the observed series y_t in a simple though reasonable manner. All the same, we should always be cognizant of the fact that the model is only an approximation to reality.

1.7 The Regression Function

The early development of time series analysis goes back to the days of Gauss who developed the method of least squares for the analysis of problems in astronomy. In such models, the effect of time was incorporated only in the systematic part $f(t)$, and not in the random part u_t . Thus it was assumed that the u_t have expectation 0, that the variance of u_t was a constant over t , and that the u_t were uncorrelated at different points in time. The systematic part $f(t)$ was a known function of time, but often involved unknown coefficients. For example, $f(t) = A + Bt$, where A and B are unknown constant; $f(t)$ is also known as the regression function.

A further analysis of $f(t)$ involves a recognition of the fact that there may be two types of sequences in time. One is a slowly moving function of time, often referred to as a trend, and is exemplified

by a polynomial in t of a fairly low degree. The other is a cyclical function of time which is exemplified by a finite sum of pairs of sine and cosine terms; this latter is called a Fourier series. For example, we could write $f(t) = \alpha \cos \lambda t + \beta \sin \lambda t$, $0 < \lambda < \pi$, where α , β , and λ are constants. We notice that the function $f(t)$ repeats itself after t has gone $2\pi/\lambda$ units of time; $2\pi/\lambda$ is therefore called the period. The reciprocal of the period $\lambda/2\pi$ is known as the frequency. The quantity $\rho = \sqrt{\alpha^2 + \beta^2}$ is called the amplitude.

Regression analysis and the theory of least squares deals with methods of inference for the unknown coefficients in the regression function $f(t)$.

Suppose now that our model for the observations y_1, y_2, \dots , is $y_t = f(t) + u_t$, where $f(t)$ is a known function of time t , and the disturbances u_t are distributed normally and independently with means 0 and variances 1. Given this information, we point out the fact that a knowledge of y_1, y_2, \dots, y_{t-1} does not give us any help in predicting y_t ; the function $f(s)$, $s > t - 1$, does not depend on y_1, y_2, \dots, y_{t-1} . However, if $f(t)$ has unknown coefficients, then the y_1, y_2, \dots, y_{t-1} can be used for estimating the unknown coefficients in $f(t)$.

1.8 Stationary Stochastic Processes

A general model in which the effect of time is represented in the random part u_t is a stationary stochastic process. For purposes of illustration, we shall consider what is known as an autoregressive process.

Suppose that y_1 has some distribution, say normal with mean 0; let the joint distribution of y_1 and y_2 be the same as the joint distribution of y_1 and $\rho y_1 + u_2$, where ρ is some constant, and u_2 has a distribution which is independent of y_1 with mean 0. We shall write $y_2 = \rho y_1 + u_2$. In general, the joint distribution of $y_1, y_2, \dots, y_{t-1}, y_t$ is the same as the joint distribution of $y_1, y_2, \dots, y_{t-1}, \rho y_{t-1} + u_t$, where u_t is distributed independently of y_1, y_2, \dots, y_{t-1} , and has mean 0. If the marginal distributions of u_2, u_3, \dots are identical (and the distribution of y_1 is specified appropriately), then (y_1, y_2, \dots, y_t) represents a segment of a stationary stochastic process, known as an autoregressive process, and

$$y_t = \rho y_{t-1} + u_t$$

is known as a stochastic difference equation of the first order.

The important notion conveyed by the above construction is that the disturbance term u_t has an effect not only on y_t , but also on the subsequent y_t 's, that is, y_{t+1}, y_{t+2}, \dots . Note the conditional expectation of y_t , given $y_{t-1}, y_{t-2}, \dots, y_1$, is

$$E(y_t | y_{t-1}, y_{t-2}, \dots, y_1) = \rho y_{t-1}.$$

Given $y_{t-1}, y_{t-2}, \dots, y_1$, our "best", in the sense of minimizing the mean square error, prediction of y_t is ρy_{t-1} . We observe here that for this model a knowledge of the earlier observations assists us in predicting y_t .

1.9 More General Models

There are many situations, particularly those involving economic time series, wherein, it may be advantageous to incorporate the effect of time in both the systematic part $f(t)$, and in the random part u_t . For example, if a series consists of a long-run movement and a seasonal variation, then $f(t)$ could be chosen to represent these features, and a u_t which represents other irregularities could be chosen to be an autoregression process.

Given a model for a time series in which the effects of time are incorporated in either the systematic part, or the random part, or both, our objectives are to estimate the unknown coefficients, test hypothesis about these coefficients, decide on the appropriate order of the process to be used, and to predict the future values of the process. In this - the first - technical report we treat statistical procedures which are concerned primarily with the systematic part $f(t)$; the random part u_t does not show the effect of time. These procedures are useful in statistical analysis, and many of them are used in subsequent approaches.

2. The Use of Regression Analysis in Times Series Analysis

Regression analysis, or what is also known as the classical least squares theory provides us with many of the techniques that are predominantly used in time series analysis. Thus it is important for us to present a brief review of the main results. The independent variables are specified functions of time, such as powers of t or trigonometric functions of t . As stated before, the random terms u_t may or may not be correlated with each other, and may or may not be normally distributed.

2.1 An Outline of the General Theory of Least Squares

Let y_1, y_2, \dots, y_T denote T observations on a time series; assume that these y 's are uncorrelated and have means

$$(2.1) \quad \bar{y}_t = \sum_{i=1}^p \beta_i z_{it} \quad , \quad t = 1, 2, \dots, T \quad ,$$

and variances

$$(2.2) \quad \sigma(y_t - \bar{y}_t)^2 = \sigma^2 \quad , \quad t = 1, 2, \dots, T \quad ,$$

where the z_{it} 's are given functions of t and are called the independent variables. The y_t 's are called the dependent variables, and the β_i 's, $i = 1, 2, \dots, p$, are p unknown coefficients; \bar{y}_t denotes the expected value of y_t .

If $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ denotes the (column) vector of the β_i 's, and if $\underline{z}_t = (z_{1t}, z_{2t}, \dots, z_{pt})'$ denotes the (column) vector of the z_{it} 's, $t = 1, 2, \dots, T$, then equation (2.1) can be written in compact form as

$$y_t = \underline{\beta}' \underline{z}_t,$$

where $\underline{\beta}'$ denotes the transpose of the vector $\underline{\beta}$.

Suppose that $\underline{\hat{b}} = (\hat{b}_1, \hat{b}_2, \dots, \hat{b}_p)'$ is any estimator of the vector $\underline{\beta}$. Then given the values y_1, y_2, \dots, y_T , our objective in least squares analysis is to find that vector \underline{b} in the class of all vectors $\underline{\hat{b}}$, such that $\sum_{t=1}^T (y_t - \underline{\hat{b}}' \underline{z}_t)^2$ is minimized. If we set $\underline{A} = \sum_{t=1}^T \underline{z}_t \underline{z}_t'$, and $\underline{C} = \sum_{t=1}^T y_t \underline{z}_t$, then it turns out that $\underline{b} = \underline{A}^{-1} \underline{C}$, where \underline{A}^{-1} denotes the inverse of \underline{A} , is the least squares estimate of $\underline{\beta}$. We, of course, assume that \underline{A} is non singular, and thus $T \geq p$.

We can verify that \underline{b} is an unbiased estimator of $\underline{\beta}$, and the covariance matrix of \underline{b} is

$$\underline{\sigma}(\underline{b} - \underline{\beta})(\underline{b} - \underline{\beta})' = \sigma^2 \underline{A}^{-1},$$

where σ^2 is the variance of the disturbance terms u_t . If $T > p$, then an unbiased estimate of σ^2 is

$$s^2 = \frac{\sum_{t=1}^T (y_t - \underline{b}' \underline{z}_t)^2}{T - p} = \frac{\sum_{t=1}^T y_t^2 - \underline{b}' \underline{A} \underline{b}}{T - p}.$$

The above results are not based on any assumptions regarding the distribution of the u_t 's (the disturbances). However, if we assume that the y_t 's are independently and normally distributed, then \underline{b} is also the maximum likelihood estimate of $\underline{\beta}$, and $s^2(T-p)/T$ is the maximum likelihood estimate of σ^2 . Furthermore, we can also show that the vector \underline{b} has a multivariate normal distribution with mean vector $\underline{\beta}$ and covariance matrix $\sigma^2 \underline{A}^{-1}$; we shall denote this distributional result by writing $\underline{b} \sim N(\underline{\beta}, \sigma^2 \underline{A}^{-1})$. Also, the quantity $(T-p)s^2/\sigma^2$ is independent of \underline{b} , and has a chi-square distribution with $T-p$ degrees of freedom, denoted by $\chi^2(T-p)$.

The main advantage of assuming that the y_t have a normal distribution is that we can test hypotheses about the β_i 's and also obtain confidence regions for them. For details about these, we refer the reader to T.W. Anderson (1971, pp. 10-11).

2.1.1 The Case of Correlated Dependent Variables

The results given above, apply when the observations y_1, y_2, \dots, y_T are not correlated. That is, when

$$\begin{aligned} E(y_t - \bar{y}_t)(y_s - \bar{y}_s) &= \sigma^2, & t = s, \\ &= 0, & t \neq s. \end{aligned}$$

Suppose now, that the dependent variables are correlated, and that the covariance matrix is known to within a constant. That is, suppose that

$$(2.3) \quad y_t = \beta' z_t, \quad t = 1, 2, \dots, T,$$

and

$$(2.4) \quad (y_t - \beta' z_t)(y_s - \beta' z_s) = \sigma^2 \psi_{ts}, \quad t, s = 1, 2, \dots, T,$$

where the ψ_{ts} are known.

We shall find it convenient to have a more complete matrix notation. Let $y = (y_1, y_2, \dots, y_T)'$, $Z = (z_1, z_2, \dots, z_T)'$, and $\Psi = [\psi_{ts}]$ denote the column vector of the y_t 's, the z_t 's and the matrix of the ψ_{ts} respectively. The matrix Z is known as the design matrix. Then

$$y = Z\beta$$

and

$$(y - Z\beta)(y - Z\beta)' = \sigma^2 \Psi.$$

The least squares estimator of β is given by

$$b = (Z'\Psi^{-1}Z)^{-1}Z'\Psi^{-1}y$$

with $b = \beta$; furthermore

$$(b - \beta)(b - \beta)' = \sigma^2 (Z'\Psi^{-1}Z)^{-1}.$$

An unbiased estimator of σ^2 is now s^2 where

$$(2.5) \quad s^2 = (\underline{y} - \underline{Z}\underline{b})' \underline{\Psi}^{-1} (\underline{y} - \underline{Z}\underline{b}) / (T - p) \quad .$$

When the observations y_t are uncorrelated and have a constant unknown variance σ^2 , $\underline{\Psi}$ will be a matrix with all its off diagonal terms equal to zero and the diagonal terms equal to 1. When this happens, the expressions given above take simpler forms and agree with the corresponding expressions of Section 2.1.

2.2 Prediction

Suppose that we wish to predict y_τ , a future observation at time τ , where $\tau > T$. If we know $\underline{\beta}$, then we know the regression function and so $\underline{\hat{y}}_\tau = \underline{\beta}' \underline{z}_\tau$ is our best predictor of y ; note that $\underline{z}_\tau = (z_{1\tau}, z_{2\tau}, \dots, z_{p\tau})'$ is assumed known. In practice, $\underline{\beta}$ will not be known and so we will have to use the observations y_1, y_2, \dots, y_T to estimate $\underline{\beta}$ and then predict y_τ using the estimated value of $\underline{\beta}$.

We can show [Anderson (1971), p. 20] that the best (in the sense of being unbiased and of minimizing the variance) estimator of $\underline{\hat{y}}_\tau$ is $\underline{b}' \underline{z}_\tau$, where \underline{b} is the least squares estimator of $\underline{\beta}$. Furthermore, the variance of our predictor $\underline{b}' \underline{z}_\tau$ is

$$\mathbb{E}(\underline{b}' \underline{z}_\tau - \underline{\beta}' \underline{z}_\tau)^2 = \sigma^2 \underline{z}_\tau' (\underline{Z}' \underline{\Psi}^{-1} \underline{Z})^{-1} \underline{z}_\tau \quad ,$$

and the mean square error of prediction is

$$\mathbb{E}(\underline{b}' \underline{z}_\tau - y_\tau)^2 = \sigma^2 (1 + \underline{z}_\tau' (\underline{Z}' \underline{\Psi}^{-1} \underline{Z})^{-1} \underline{z}_\tau) \quad .$$

When the observations y_t are uncorrelated and have a constant unknown variance σ^2 , the mean square error of prediction becomes $\sigma^2(1 + \underline{z}_t' \underline{A}^{-1} \underline{z}_t)$, where $\underline{A} = \underline{Z}' \underline{Z}$.

If we assume that the y_t 's are distributed normally then a prediction interval for y_t can be obtained using the fact [cf. Anderson (1971), p. 21] that

$$\frac{\underline{b}' \underline{z}_t - y_t}{s \sqrt{1 + \underline{z}_t' (\underline{Z}' \underline{\Psi}^{-1} \underline{Z})^{-1} \underline{z}_t}}$$

has a Student t-distribution with $T - p$ degrees of freedom. Thus a prediction interval for y_t with confidence $1 - \epsilon$ is given by

$$(2.6) \quad \underline{b}' \underline{z}_t \pm t_{T-p}(\epsilon) s \sqrt{\underline{z}_t' (\underline{Z}' \underline{\Psi}^{-1} \underline{Z})^{-1} \underline{z}_t + 1}$$

where s is obtained via equation (2.5) and $t_{T-p}(\epsilon)$ is the number such that the probability of a Student t-distribution with $(T - p)$ degrees of freedom between $\pm t_{T-p}(\epsilon)$ is $1 - \epsilon$.

When we cannot assume normality of the observations, the above procedures can still be justified for large samples on the basis of asymptotic theory (Anderson (1971), p. 23).

2.3 An Illustrative Example

We shall illustrate our use of the methods discussed thus far by considering the following example involving some real-life data. We would like to emphasize that our main objective in presenting this

example is to illustrate the mechanics of using the methodology. Our goal is not to solve realistically the practical problems posed by the example.

In Table 2.1, we show some data (taken from Beals, 1972, p. 297) on the demand for money, y_t , for year t , in Jamaican dollars, for the years 1961 through 1970. Also shown are the data on the gross national product Z_{1t} , and the treasury bill rate Z_{2t} . For the

Table 2.1

Money Stock, GNP, and Treasury Bill Rate
for Jamaica, 1961-1970

Year	Money Stock in millions of Jamaican dollars	GNP in millions of Jamaican dollars	Treasury Bill Rate %
1961	57.2	485.5	4.01
1962	55.6	506.9	4.93
1963	54.4	540.2	4.18
1964	62.6	589.0	3.27
1965	67.2	637.4	4.35
1966	66.4	692.7	4.44
1967	70.4	745.3	4.40
1968	82.8	821.6	4.71
1969	102.4	906.3	3.33
1970	115.0	981.3	3.90

purposes of this example, following Beals, we shall assume that the demand for money (money stock) is a linear function of the gross national product and the treasury bill rate. That is

$$y_t = \beta_0 Z_{0t} + \beta_1 Z_{1t} + \beta_2 Z_{2t} + u_t, \quad t = 1961, \dots, 1970,$$

where $Z_{0t} = 1$ for all values of t , and β_0 , β_1 , and β_2 are unknown constants.

We shall treat the random disturbance terms as being normally distributed with mean zero and an unknown constant variance σ^2 .

A graph of the data of Table 2.1 is shown in Figure 2.1. An inspection of Figure 2.1 reveals the fact that whereas the money stock and the gross national product increase with time, the treasury bill rates do not appear to reflect an upward movement.

In vector notation, $\beta = (\beta_0, \beta_1, \beta_2)$ and $Z_t = (Z_{0t}, Z_{1t}, Z_{2t})'$; thus, our model for the money stock series, is $y_t = \beta' Z_t$, with $\sum_{i=0}^2 \beta_i Z_{it}$ and $\sum (y_t - \hat{y}_t)^2 = \sigma^2$. Our data on the money stock is $y = (57.2, 55.6, \dots, 115.0)$, and our design matrix is $Z = (Z_{1961}, \dots, Z_{1970})'$, where, for example, $Z_{1961} = (1, 485.5, 4.01)'$.

The least squares estimate of β is $\hat{\beta} = (b_0, b_1, b_2) = (Z'Z)^{-1} Z'y = (17.059, .1114, -4.961)'$. An unbiased estimate of σ^2 is s^2 , where $s^2 = (y'y - \hat{\beta}' A \hat{\beta}) / (T - p) = 263.3714 / 7 = 37.624$, where $A = Z'Z$. In order to be able to test the significance of the estimated coefficients $\hat{\beta}$, we need to first compute their standard errors, which can be obtained from $s^2 A^{-1}$. These turn out to be 6.134, .0123, and 3.893, respectively.

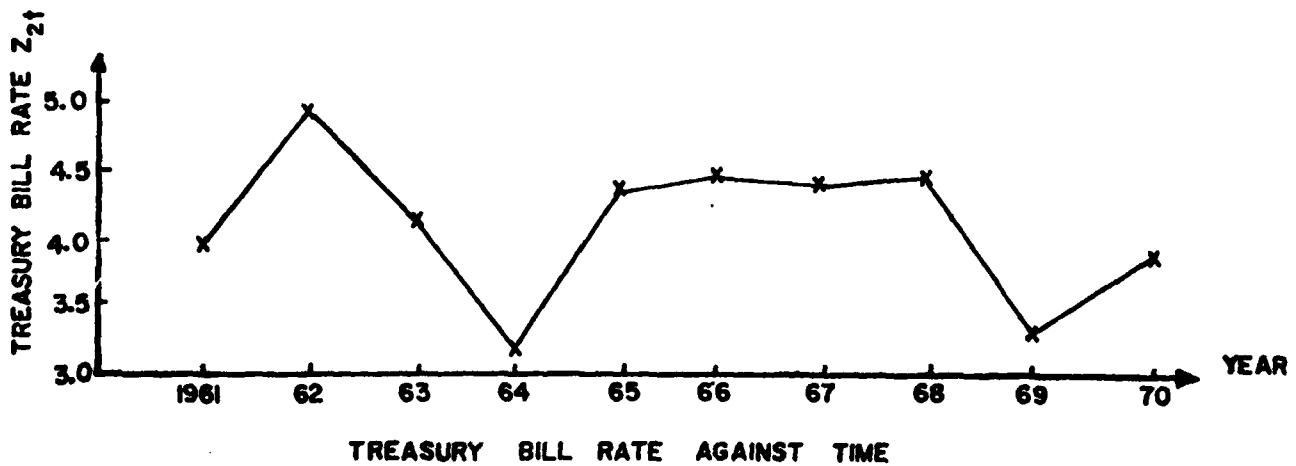
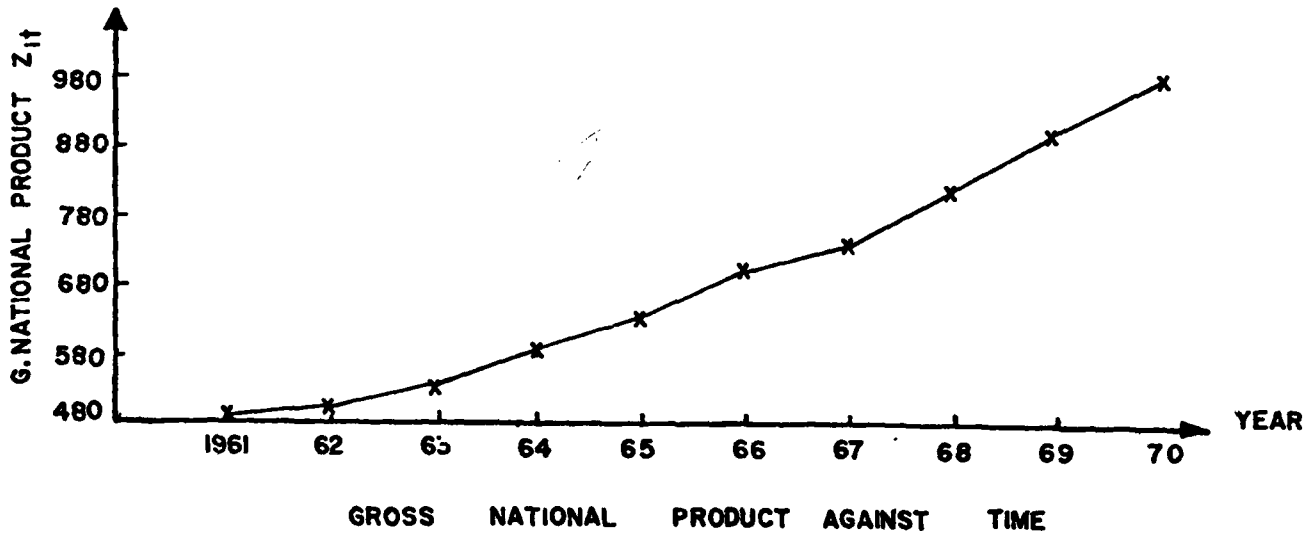
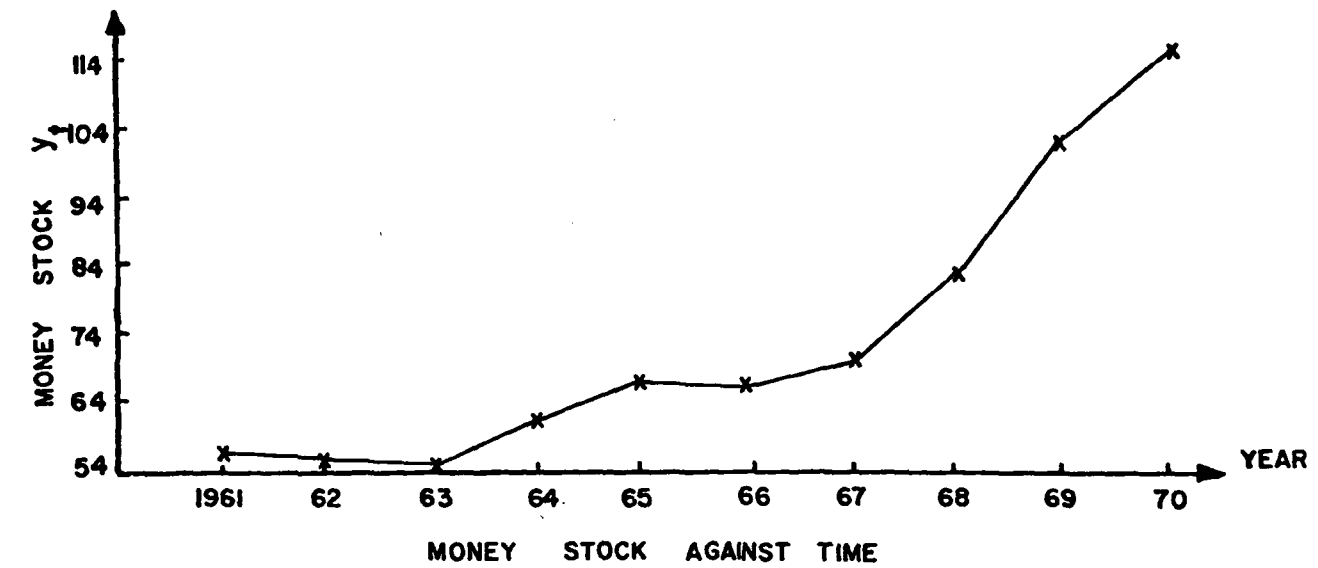


Figure 2.1. Money stock, gross national product and treasury bill rates for Jamaica.

To test for the significance of these coefficients, we compute their t-statistic values, $17.058/6.134 = 2.7809$, $.1114/.0123 = 9.057$, and $-4.96/3.893 = -1.27$. Since the length of the series is small, the significance level should be large, say 10% or so. Economic theory says that both the GNP and the interest rates affect money supply, thus a test of significance of the coefficients β_1 and β_2 should be one sided. The fact that β_2 has a negative coefficient makes economic sense, since an increase in the treasury bill rate will tend to lower the money stock. The critical value of the t-statistic with 7 degrees of freedom, for a one-sided 10% level of significant test is 1.415. Thus the coefficients β_0 and β_1 are clearly significant, whereas the coefficient β_2 is nearly significant. In view of the above, together with the fact that interest rates are known to affect money supply, the coefficient β_2 is retained in the model. A summary of the pertinent test statistics is given in Table 2.2.

To continue with our analysis, we compute the "fitted values" \hat{y}_t , where

$$\hat{y}_t = 17.059 + .1114 Z_{1t} - 4.961 Z_{2t} \quad , \quad t = 1961, \dots, 1970 \quad .$$

These values are shown in column 3 of Table 2.3. The "residuals" $y_t - \hat{y}_t$ are shown in column 4 of Table 2.3; they are plotted in Figure 2.2a. In Figure 2.2b we plot \hat{y}_t versus y_t , and note that the two compare well. The residual sum of squares, $\sum_{t=1}^{10} (y_t - \hat{y}_t)^2 = 263.371$;

Table 2.2

Summary Statistics for Regression Analysis of Data on
Money Stock, GNP and Treasury Bill Rates for Jamaica

Variable	Coefficient	Standard Error	t-statistic
Constant	17.0585	6.134	2.781
GNP	.1114	.0123	9.057
Treasury Bill Rate	-4.9611	3.893	-1.27

$$\bar{y} = \sum_{t=1}^{10} y_t / 10 = 73.400 \quad ; \quad \sum_{t=1}^{10} (y_t - \bar{y})^2 = 3813.28$$

$$\bar{z}_{1t} = \sum_{t=1}^{10} z_{1t} / 10 = 690.619 \quad ; \quad \bar{z}_{2t} = \sum_{t=1}^{10} z_{2t} / 10 = 4.152$$

$$R^2 = .931 \quad .$$

Table 2.3

A Comparison of Actual Versus Fitted Values of the Money Stock
for Jamaica, 1961-1970

Year	Actual Values of Money Stock y_t	Fitted Values \hat{y}_t	Residuals $y_t - \hat{y}_t$
1961	57.2	51.25	5.95
1962	55.6	49.07	6.53
1963	54.4	56.50	-2.1
1964	62.6	66.45	-3.85
1965	67.2	66.48	.72
1966	66.4	72.20	-5.8
1967	70.4	78.26	-7.86
1968	82.8	85.22	-2.42
1969	102.4	101.50	.9
1970	115.0	107.03	7.97

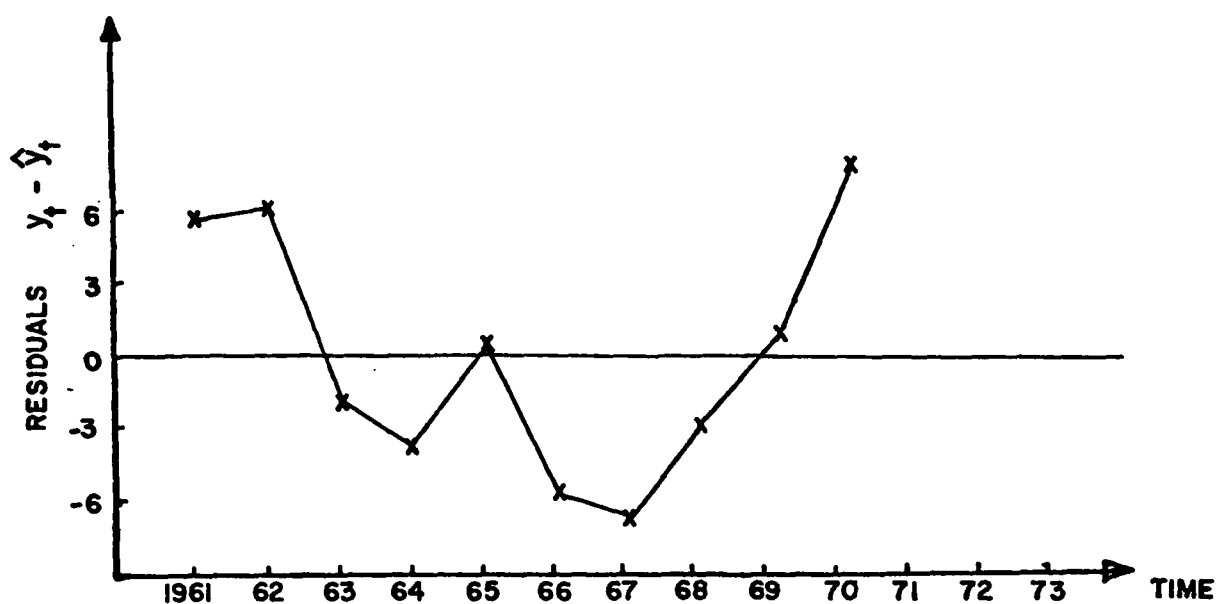


Figure 2.2(a). Residuals ($y_t - \hat{y}_t$) against time for data on money stock for Jamaica.

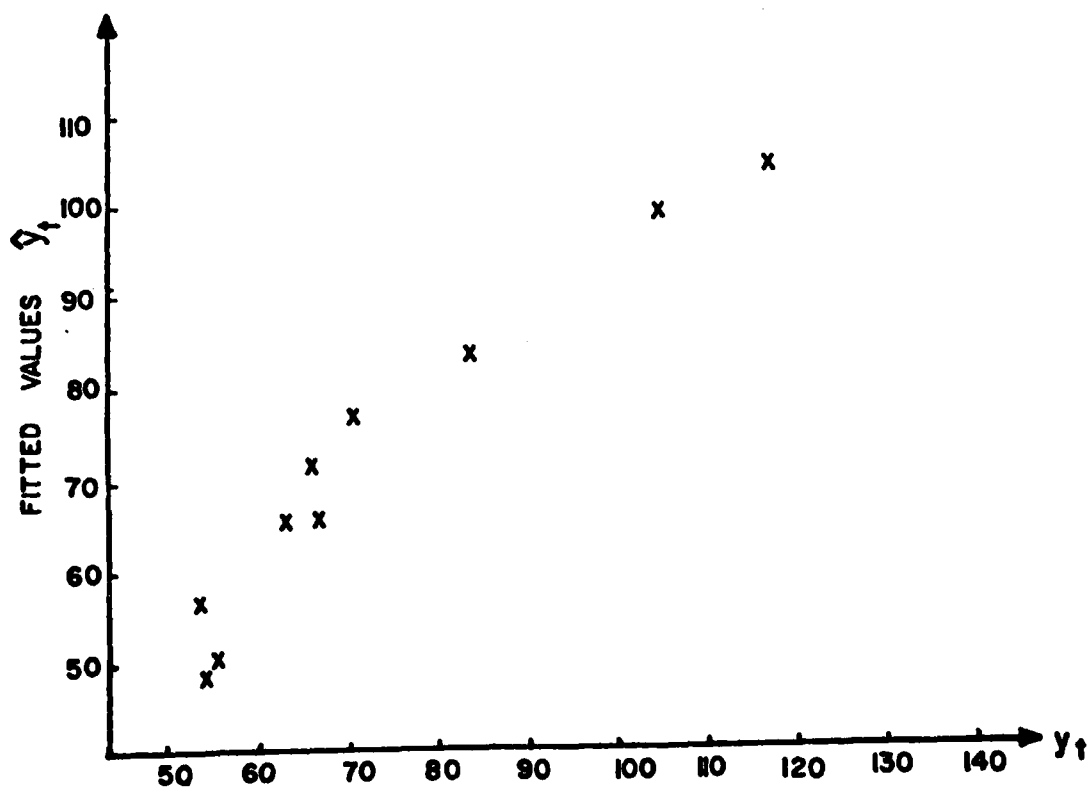


Figure 2.2(b). Fitted values \hat{y}_t against actual values y_t for data on money stock in Jamaica.

this quantity when subtracted from the original sum of squares $\sum_{t=1}^{10} (y_t - \bar{y})^2 = 3813.28$, gives us 3549.909, which is the sum of squares due to the regression of y_t on z_{1t} and z_{2t} . The ratio $3549.909/3813.28 = .9309$ is the R^2 value, and this a measure of how well the chosen model explains the variability of y_t .

Conclusion

The purpose of this analysis is an explanation of the behavior of the money supply, rather than its prediction, since in practice we do not know in advance the bill rates. Our conclusion is that the GNP and the bill rates do affect the money stock, in a manner indicated by the model.

3. Trends and Smoothing

A trend is a broad movement, either upwards or downwards, in a time series. Trends are generally of great interest to businessmen and economists. Given a time series, we may wish to infer the presence or the absence of a trend. Often, we may know of a physical or an economic principle which guides us in the specification of the functional form of the trend. For example, in reliability theory it is common to assume that the failure rate of items which age with time increases as either a linear or a quadratic function of time. [See Mann, Schafer, Singpurwalla (1975).] Thus, the estimated failure rate which constitutes a natural time series [Singpurwalla (1975)] should contain a term which accounts for this trend. However, there are many real life situations wherein we do not have an underlying reason for specifying the general form of the trend, and so we may wish to approximate it by a polynomial of a very low degree.

When we use such polynomial functions to describe the trend, we should bear in mind that these functions are approximations to an unknown function of time. The true unknown function may be much more complicated than the approximating polynomial. Consequently, we cannot give any real physical meaning to the coefficients of the polynomial. Furthermore, the polynomial can be used for interpolation only; extrapolations must be made with great caution.

3.1 Polynomial Trends

Following Section 2, let us assume that an observation y_t , $t = 1, 2, \dots, T$, is the sum of $f(t)$ which is a trend in t , and an error (disturbance) term u_t , where $\sum u_t = 0$, $\sum u_t^2 = \sigma^2$, and $\sum u_t u_s = 0$, $t \neq s$. Suppose further that the trend is a polynomial in t of degree p , where $p < T$, that is,

$$(3.1) \quad y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_p t^p + u_t .$$

In practice, we do not know the β 's nor do we know the value of p , and our objective is to choose the smallest value of p which is consistent with the observed y_t and (3.1). A point of view that we may take here is that in fitting a polynomial trend, or for that matter any other regressive function, our goal is a reduction of the data. If we set $t^{i-1} = z_{it}$, $i = 1, 2, \dots, p+1$, then the condition given by (2.1) and (2.2) are satisfied by the above model, and for any specified value of p , we can use the method of least squares for estimating the β 's. However, a considerable amount of simplification, both computational and statistical, can result if we use orthogonal polynomials.

3.1.1 Orthogonal Polynomials

We say that column i of a given matrix is orthogonal to column j of the matrix if the sum of their corresponding cross products is zero. Our aim is to have the columns of the design matrix Z orthogonal to each other. This is achieved by transforming the independent variables

$1, t, t^2, \dots, t^p$ (all of which are powers of t) to a set of what are known as orthogonal independent variables $\phi_{0T}(t), \phi_{1T}(t), \dots, \phi_{pT}(t)$, where

$$\sum_{t=1}^T \phi_{iT}(t) \phi_{kT}(t) = 0, \quad i \neq k, \quad i, k = 0, 1, \dots, p.$$

For purpose of illustration, we have $\phi_{0T}(t) = 1$, a polynomial in t of order 0, $\phi_{1T}(t) = t - (1/2)(T+1)$, a polynomial in t of order 1, $\phi_{2T}(t) = t^2 - (T+1)t + (T+1)(T+2)/6$, a polynomial in t of order 2, and so on. The coefficient of t^k in the orthogonal polynomial of degree k is often taken to be 1.

Fisher and Yates (1963) give the values of orthogonal polynomials up to degree 5 for T up to 75. Another source for obtaining these polynomials is Biometrika Tables for Statisticians, Vol. 1. (In these tables the coefficient of t^k is taken so that the values of the orthogonal polynomial of order p are integers; examples are given later.) Of course, tables are not needed to carry out these procedures with a computer because the program to construct orthogonal polynomials is easily available.

In terms of these orthogonal polynomials, (3.1) can be written as

$$y_t = \gamma_0 \phi_{0T}(t) + \gamma_1 \phi_{1T}(t) + \dots + \gamma_p \phi_{pT}(t) + u_t,$$

where $\gamma_0, \dots, \gamma_p$ are a new set of constants which are linearly related to the original set of constants β_0, \dots, β_p ; in fact, the last coefficient γ_p equals β_p .

For the model given by equation (3.1), the design matrix Z is

$$\underline{Z} = \begin{bmatrix} \phi_{0T}(1) & \phi_{1T}(1) & \dots & \phi_{pT}(1) \\ \phi_{0T}(2) & \phi_{1T}(2) & \dots & \phi_{pT}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{0T}(T) & \phi_{1T}(T) & \dots & \phi_{pT}(T) \end{bmatrix}$$

If we write $a_{kk} = \sum_{t=1}^T (\phi_{kT}(t))^2$, then

$$\underline{Z}'\underline{Z} = \begin{bmatrix} a_{00} & 0 & \dots & 0 \\ 0 & a_{11} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{pp} \end{bmatrix},$$

the orthogonality property making the off diagonal terms equal to zero.

If we set $\underline{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_p)'$, then the least squares estimator of $\underline{\gamma}$ is $\hat{\underline{\gamma}} = (\underline{Z}'\underline{Z})^{-1}\underline{Z}'\underline{y}$; thus an unbiased estimator of γ_k is

$$\hat{\gamma}_k = \frac{\sum_{t=1}^T y_t \phi_{kT}(t)}{\sum_{t=1}^T \phi_{kT}^2(t)}, \quad k = 0, 1, \dots, p.$$

The covariance matrix of $\hat{\underline{\gamma}}$ is $(\underline{Z}'\underline{Z})^{-1}\sigma^2$, from which we see that

$$\text{Var}(\hat{\gamma}_k) = \frac{\sigma^2}{a_{kk}}, \quad k = 0, 1, \dots, p;$$

an estimator of σ^2 is

$$s^2 = \frac{\sum_{t=1}^T (y_t - \sum_{k=0}^p \hat{\gamma}_k \phi_{kT}(t))^2}{T - p - 1} = \frac{\sum_{t=1}^T y_t^2 - \sum_{k=0}^p a_{kk} \hat{\gamma}_k^2}{T - p - 1} .$$

We remark that the above particularly simple forms of our estimators is due to our orthogonalization procedure. Another advantage of orthogonalization is that the estimates $\hat{\gamma}_k$ are uncorrelated, since the off-diagonal elements of $(Z'Z)^{-1}$ are 0.

If we assume that the y_t 's are normally distributed, then $\hat{\gamma}_0, \dots, \hat{\gamma}_p$ are independently and normally distributed, and $(T-p-1)s^2/\sigma^2$ is distributed as χ^2 with $T-p-1$ degrees of freedom independently of $\hat{\gamma}_0, \dots, \hat{\gamma}_p$. We can therefore test the null hypothesis that $\gamma_k = 0$ versus the alternative that $\gamma_k \neq 0$ at significance level α by using the Student t-distribution with $T-p-1$ degrees of freedom. That is, we reject the hypothesis that $\gamma_k = 0$ whenever

$$(3.3) \quad \frac{|\hat{\gamma}_k| \sqrt{a_{kk}}}{s} > t_{T-p-1}(\alpha) ,$$

where $t_{T-p-1}(\alpha)$ is the two-sided α -significance point of the Student t-distribution with $T-p-1$ degrees of freedom.

3.1.2 Determining the Degree of Polynomial Trend

Suppose, for the moment, that we have reason to believe that the degree p of the polynomial trend is at most a specified q . Then a test

of the null hypothesis that the polynomial trend is of degree less than k , given that its degree is at most k , for any $k \leq q$, is a test of the null hypothesis that $\gamma_k = 0$ against the alternative that $\gamma_k \neq 0$; this test can be performed using the procedure given by (3.3).

In practice, we do not know in advance the value of p that we should use. Our inclination is to use as low a value of p as is possible, so that our curve $f(t)$ is smooth and economical. However, a disadvantage of choosing too low a value of p is that a bias is introduced in our estimate of the trend. To overcome this dilemma, suppose that we have some a priori information which leads us to believe that the lowest value that p can take is m , where m could be zero, and that the maximum value which p can take is some value q . We are now confronted with the problem of deciding whether the degree of our polynomial is $m, m + 1, \dots, q - 1$, or q . A natural strategy would be to work forward by starting off by choosing $p = m + 1$, and then testing the null hypothesis that $\gamma_{m+1} = 0$ using (3.3); if this hypothesis is rejected, then we test the hypothesis that $\gamma_{m+2} = 0$, and continue in this manner until some hypothesis is accepted or until $\gamma_q = 0$ has been rejected. However, as is discussed on p. 42 of Anderson (1971), this approach could lead us to an erroneous decision. A better procedure is to proceed backward by starting off by choosing $p = q$, and

then testing the null hypothesis that $\gamma_q = 0$ using (3.3); if this hypothesis is accepted, then we choose $p = q - 1$, and test the null hypothesis that $\gamma_{q-1} = 0$ using equation (3.3) with the value of s being recomputed. Note that when we use orthogonal polynomials our estimates of the coefficients do not change when we go from $p = q$ to $p = q - 1$; this is another advantage of using orthogonal polynomials. We continue in this manner until some hypothesis is rejected or until $\gamma_{m+1} = 0$ has been accepted. If in the above (backward) procedure, suppose that γ_{q-j} , $j = 1, 2, \dots, q - m - 1$, is the first hypothesis to be rejected. Then our conclusion would be that the polynomial trend $f(t)$ is of degree $q - j$; we do not proceed to test if $\gamma_{q-j-1} = 0$.

3.1.3 An Example Illustrating the Fitting of a Polynomial Trend

In Table 3.1 (taken from T.W. Anderson (1971), p. 44), we show y_t the quantity of meat consumed per year per person in the United States from 1919 to 1941; thus $T = 23$. A plot of y_t versus t is shown in Figure 3.1. Our aim in the analysis presented below is twofold:

- (i) to illustrate the methodology for fitting polynomial trends, and
- (ii) to attempt to make some comments on the nature of the trend, if any, based on the fitted polynomial.

Table 3.1

Annual Consumption of Meat in the United States,
from 1919 - 1941

Time Period 1918 + t	Annual Consumption of Meat y_t	Fitted Values \hat{y}_t Using A Third Degree Polynomial
1	171.5	165.805
2	167.0	169.456
3	164.5	171.927
4	169.3	173.350
5	179.4	173.859
6	179.2	173.585
7	172.6	172.662
8	170.5	171.223
9	168.6	169.399
10	164.7	167.325
11	163.0	165.132
12	162.1	162.954
13	160.2	160.923
14	161.2	159.172
15	165.8	157.833
16	163.5	157.040
17	146.7	156.925
18	160.2	157.620
19	156.8	159.260
20	156.8	161.975
21	165.4	165.900
22	174.7	171.167
23	178.7	177.908

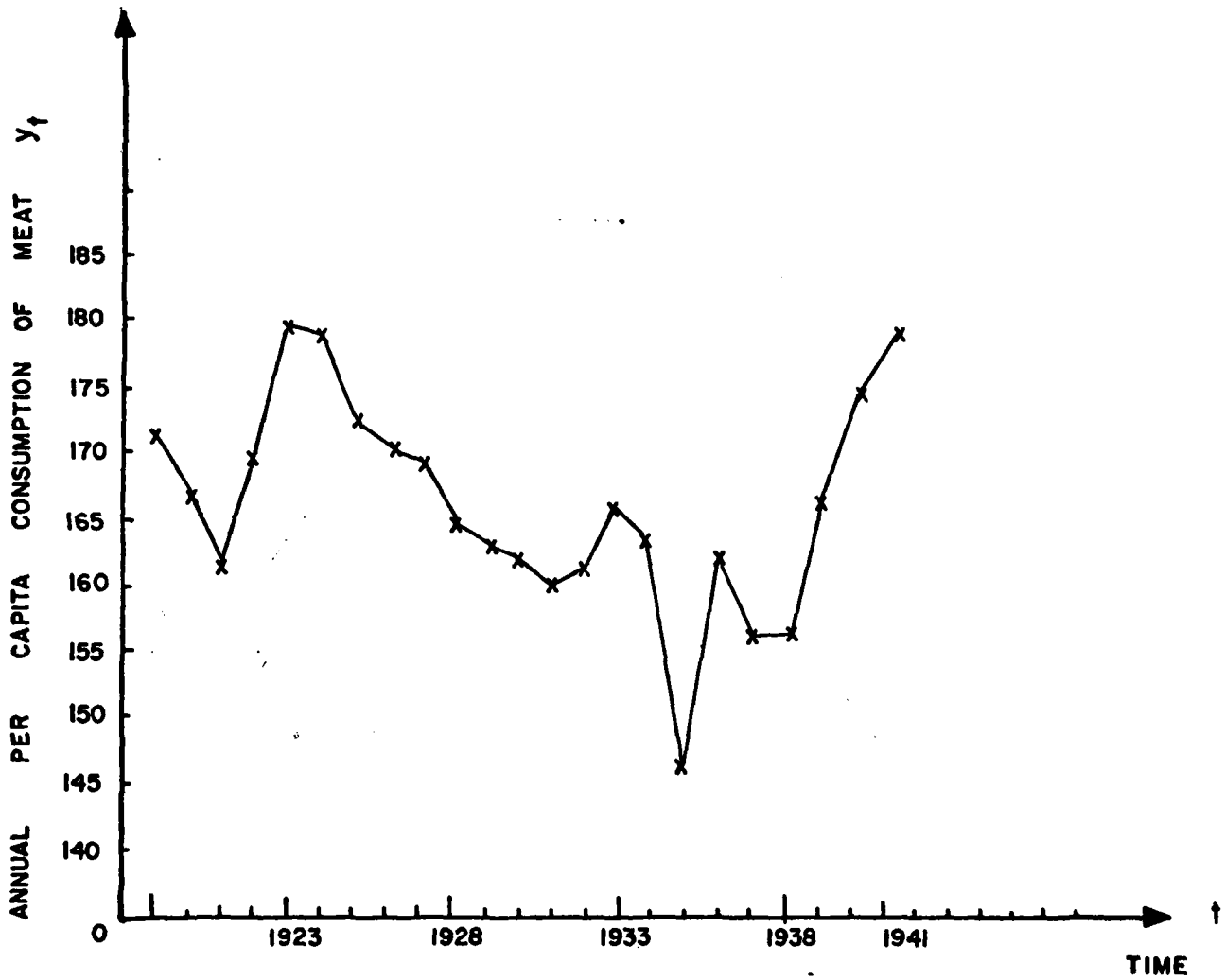


Figure 3.1. Annual per capita consumption of meat in the United States, 1919-1941.

We treat the y_t 's as being independent and normally distributed with an unknown variance σ^2 . An inspection of Figure 3.1 suggests a cyclical pattern; motivated by this, we shall suppose that the trend (if any) can be described by a polynomial in t of degree at least 3 and at most 5. A polynomial of degree 3 being a cubic function, is suitable for describing the cyclical pattern in the data.

We start off by fitting a polynomial of degree 5 using the model

$$y_t = \gamma_0 \phi_{0,23}(t) + \gamma_1 \phi_{1,23}(t) + \dots + \gamma_5 \phi_{5,23}(t) + u_t, \\ t = 1, 2, \dots, 23,$$

where $\gamma_0, \dots, \gamma_5$ are the unknown coefficients. Our design matrix is

$$\tilde{Z} = \begin{bmatrix} \phi_{0,23}(1) & \phi_{1,23}(1) & \dots & \phi_{5,23}(1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{0,23}(23) & \phi_{1,23}(23) & \dots & \phi_{5,23}(23) \end{bmatrix}$$

with the entries $\phi_{k,23}(t)$, $t = 1, \dots, 23$, $k = 0, \dots, 5$, given in Table 3.2, are taken from Biometrika Tables for Statisticians, Vol. 1, p. 215. The entries in Table 3.2 are the values of the orthogonal polynomials with leading coefficients $\lambda_0, \dots, \lambda_5$ chosen to make the values integers. For example, when $T = 23$, and when $x(t) = t - \frac{1}{2}(T+1)$,

$$\phi_{0,T}(t) = \lambda_0, \quad \phi_{1,T}(t) = \lambda_1 x(t), \\ \phi_{2,T}(t) = \lambda_2 \{x^2(t) - \frac{1}{12}(T^2 - 1)\},$$

$$\phi_{3,T}(t) = \lambda_3 \{x^3(t) - \frac{1}{20} (3T^2 - 7)x(t)\} ,$$

and so on, where the values $\lambda_0, \dots, \lambda_5$ are also given in Table 3.2. The advantage of using the above formulas is that the orthogonal polynomials are displayed in integers, with the result that rounding errors are avoided.

Let $\gamma = (\gamma_0, \dots, \gamma_5)'$; then $\hat{\gamma}$ the least squares estimator of γ is $\hat{\gamma} = (Z'Z)^{-1}Z'Y$; thus an unbiased estimator of γ_k is

$$\hat{\gamma}_k = \sum_{t=1}^{23} y_t \phi_{k,23}(t) / \sum_{t=1}^{23} \phi_{k,23}^2(t) , \quad k = 0, \dots, 5 .$$

We find $\hat{\gamma} = (166.191, -.379051, .073574, .132745, .001314, .000052)$ and the estimate of σ^2 is

$$s^2 = \sum_{t=1}^{23} (y_t - \sum_{k=0}^5 \hat{\gamma}_k \phi_{k,23}(t))^2 / (23 - 5 - 1) = 25.303 .$$

In order to see if the polynomial trend has degree less than 5, given that it has degree at most 5 we compute $|\hat{\gamma}_5| \sqrt{a_{55}} / \sqrt{s^2}$, where $a_{55} = \sum_{t=1}^{23} (\phi_{5,23}(t))^2 = 340860$; thus, we have $(.000052)(583.832) / \sqrt{25.303} = .00608$. Since $t_{17}(.05) = 2.11$, we accept the null hypothesis that $\gamma_5 = 0$, and conclude that the polynomial trend can be described by a polynomial of order less than 5. To test the hypothesis that $\gamma_4 = 0$, we compute $s^2 = \sum_{t=1}^{23} (y_t - \sum_{k=0}^4 \hat{\gamma}_k \phi_{k,23}(t))^2 / (23 - 4 - 1) = 23.8973$, and now $|\hat{\gamma}_4| \sqrt{a_{44}} / \sqrt{23.8973}$ turns out to be .9738; since $t_{16}(.05) = 2.12$,

Table 3.2

Orthogonal Polynomials for Fitting a Polynomial
of Degree 5 to 23 Observations

t	$\phi_{0,23}(t)$	$\phi_{1,23}(t)$	$\phi_{2,23}(t)$	$\phi_{3,23}(t)$	$\phi_{4,23}(t)$	$\phi_{5,23}(t)$
1	1	-11	77	-77	1463	-209
2	1	-10	56	-35	133	76
3	1	-9	37	-3	-627	171
4	1	-8	20	20	-950	152
5	1	-7	5	35	-955	77
6	1	-6	-8	43	-747	-12
7	1	-5	-19	45	-417	-87
8	1	-4	-28	42	-42	-132
9	1	-3	-35	35	315	-141
10	1	-2	-40	25	605	-116
11	1	-1	-43	13	793	-65
12	1	0	-44	0	858	0
13	1	1	-43	-13	793	65
14	1	2	-40	-25	605	116
15	1	3	-35	-35	315	141
16	1	4	-28	-42	-42	132
17	1	5	-19	-45	-417	87
18	1	6	-8	-43	-747	12
19	1	7	5	-35	-955	-77
20	1	8	20	-20	-950	-152
21	1	9	37	3	-627	-171
22	1	10	56	35	133	-76
23	1	11	77	77	1463	209
λ_k	1	1	1	$\frac{1}{6}$	$\frac{7}{12}$	$\frac{1}{60}$
$\sum_{t=1}^{23} \phi_{k,23}^2(t)$	23	1012	35,420	32,890	13,123,110	340,860

we accept the null hypothesis that $\gamma_4 = 0$. We now carry out the computations to test if $\gamma_3 = 0$; the appropriate test statistic works out to be 4.93, which because of $t_{15}(.05)$ being equal to 2.131, that $\gamma_3 = 0$, and conclude that the trend can be described by a polynomial of order 3. In Table 3.2.1, we summarize our computations with the orthogonal polynomials.

Our fitted polynomial is therefore

$$\begin{aligned}\hat{y}_t &= 166.191\phi_{0,23}(t) - .379\phi_{1,23}(t) + .074\phi_{2,23}(t) + .133\phi_{3,23}(t) , \\ t &= 1, \dots, 23 \quad . \\ &= 166.191 - .379(t - 12) + .074\{(t - 12)^2 - \frac{1}{12} \times 528\} \\ &\quad + \frac{.133}{6} \{(t - 12)^3 - \frac{1}{20}(3 \times 529 - 7)(t - 12)\} \\ &= 160.849 + 5.670t - .724t^2 + .022t^3 \quad .\end{aligned}$$

A graph of the fitted polynomial together with the actual values of y_t is shown in Figure 3.2; the values of \hat{y}_t are given in column 3 of Table 3.1. It will be observed that Figure 3.2 gives a good fit, most of the points being close to the curve. We interpret the curve as the expected or normal consumption of meat if it were not affected by year-to-year irregularities. The fitted third-degree polynomial cannot be good for prediction - at least not very far in the future. Far to the right of this data, this polynomial increases and with an increasing slope; even without the effect of war, it does not seem reasonable that per capita meat consumption will increase indefinitely at an increasing rate of increase.

Table 3.2.1

Computations with Orthogonal Polynomials for the
Example of Meat Consumption

Degree of Polynomial k	$a_{kk} = \sum_{t=1}^{23} \phi_{k,23}^2(t)$	$\sum_{t=1}^{23} y_t \phi_{k,23}(t)$	Coefficient \hat{y}_k	Square $\hat{y}_k^2 a_{kk}$	Residual Sum of Squares	Mean Residual Sum of Squares	Ratio t^2
0	23		166.191		1369.538	62.251	
1	1,012	-383.6	-.379,051	145.404	1224.413	58.305	2.494
2	35,420	2,606.0	.073,574	191.735	1032.400	51.620	3.714
3	32,890	4,366.0	.132,745	579.567	452.833	23.833	24.318
4	13,123,110	17,252.4	.001,314	22.681	430.152	23.897	.949
5	340,860	17.8	.000,052	.000,929	430.151	25.303	.000,037

$$\sum_{t=1}^{23} y_t = 3822.4, \quad \sum_{t=1}^{23} y_t^2 = 636,619.18$$

ANNUAL PER CAPITA CONSUMPTION OF MEAT, y_t

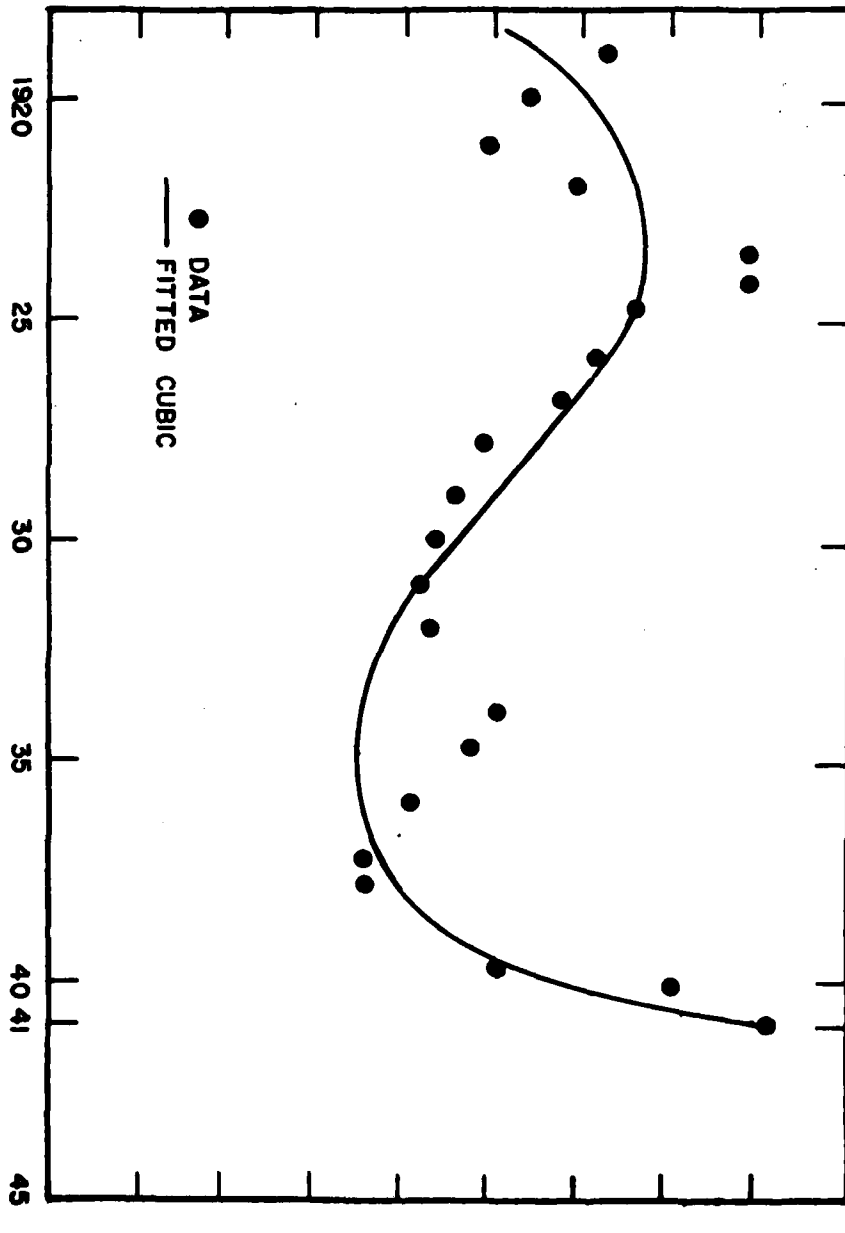


Figure 3.2. Third degree polynomial fit to data on annual consumption of meat.

Instead of using cubic terms to describe the cyclical pattern in the data, a more effective way might be to use Fourier series (sums of sine and cosine terms). This topic will be considered later on. In Example 2 of Section 4, we will reconsider this data and illustrate how the technique of smoothing can be used to reduce its variability.

3.2 Smoothing

In order to estimate the trend at a given point in time, it may be more meaningful to consider only those observations which are in the neighborhood of the time point rather than all the observations, as was done in Section 3.1. A procedure which accomplished this is known as smoothing; here the trend at a given point in time is the weighted average of the observations in the vicinity of that point. If we smooth over all the points in time (except the first and last few), then an irregular graph of the observed points will be replaced by a smooth graph. Specifically, given the time series, y_1, y_2, \dots, y_T , an estimate of the trend at time t is given by

$$y_t^* = \sum_{s=-m}^{s=m} c_s y_{t+s} \quad , \quad t = m + 1, \dots, T - m \quad ,$$

where m is some suitable constant, and the c_s 's are weight functions which sum to 1; that is $\sum_{s=-m}^m c_s = 1$. The observed sequence $\{y_t^*\}$,

$t = m + 1, \dots, T - m$, is called a moving average^{1/} of the original sequence $\{y_t\}$. Since we have assumed that $y_t = f(t) + u_t$, it follows that

$$\begin{aligned} y_t^* &= \sum_{s=-m}^m c_s y_{t+s} = \sum_{s=-m}^m (c_s f(t+s) + c_s u_{t+s}) \\ &= \sum_{s=-m}^m c_s f(t+s) + u_t^* , \end{aligned}$$

where $u_t^* = \sum_{s=-m}^m c_s u_{t+s}$.

Since $\mathbb{E}u_t = 0$, $\mathbb{E}u_t^2 = \sigma^2$ and $\mathbb{E}u_t u_s = 0$, $t \neq s$, $\mathbb{E}u_t^{*2} = \sigma^2 \sum_{s=-m}^m c_s^2$; we should choose the c_s 's in such a manner that $\mathbb{E}u_t^{*2}$ is considerably smaller than $\mathbb{E}u_t^2$. Furthermore, since $\mathbb{E}y_t^* = \sum_{s=-m}^m c_s f(t+s)$, $\mathbb{E}y_t^* \neq f(t)$; and so unless the values $f(t+s)$, $s = 1, \dots, m$ are all close to $f(t)$ (that is, the trend does not change rapidly), the smoothing will introduce some bias. In general, the smoothed sequence $\{y_t^*\}$ has a smaller variance than the original sequence $\{y_t\}$, but is biased. Another important consequence of smoothing is that the successive terms in the smoothed sequence are correlated, even though the original sequence was not. Specifically, for any $h \geq 0$,

$$\begin{aligned} \mathbb{E}u_t^* u_{t+h}^* &= \sum_{s=-m}^m \sum_{r=-m}^m c_s c_r \mathbb{E}u_{t+s} u_{t+h+r} \\ &= \sigma^2 \sum_{s=-m+h}^m c_s c_{s-h} , \quad h = 0, 1, \dots, 2m \\ &= 0 , \quad h = 2m + 1, \dots, . \end{aligned}$$

^{1/}The moving average process, to be introduced later, is mathematically equivalent to the sequence $\{y_t^*\}$, but usually the trend term is absent and it is the random sequence $\{y_t^*\}$ that is relevant.

As an example, suppose that we choose $c_s = 1/(2m + 1)$; that is, we take the arithmetic average of the $2m + 1$ points. Then

$$y_t^* = \frac{1}{2m + 1} \sum_{s=-m}^m f(t + s) + \frac{1}{2m + 1} \sum_{s=-m}^m u_{t+s} ,$$

where $\sigma_{u_t^*}^2 = \sigma^2/(2m + 1)$, which is less than σ^2 , and

$$\begin{aligned} \sigma_{u_t^* u_{t+h}^*} &= \frac{2m + 1 - h}{(2m + 1)^2} \sigma^2 , \quad \text{for } h = 0, 1, \dots, 2m , \\ &= 0 , \quad \text{otherwise .} \end{aligned}$$

Clearly, the variance of u_t^* can be reduced by choosing m large, but then $|y_t^* - f(t)|$ may also increase.

3.2.1 The Theory Underlying the Smoothing Procedure and Method for Obtaining the Smoothing Coefficients

The notion underlying the smoothing procedure is that instead of fitting a polynomial of degree p to the entire set of data (as was done in Section 3.1), we fit a polynomial of degree p to $2m + 1$ successive values, and then use this polynomial to estimate the trend at the middle value. Suppose that we consider the $2m + 1$ time points around t , say $t, t \pm 1, t \pm 2, \dots, t \pm m$, and suppose that the trend at these time points $f(t + s)$, $s = 0, \pm 1, \dots, \pm m$, can be approximated by the polynomial

$$\begin{aligned} \beta_0 + \beta_1 s + \beta_2 s^2 + \dots + \beta_p s^p , \\ s = 0, \pm 1, \pm 2, \dots, \pm m . \end{aligned}$$

Then, our trend at time t , $f(t)$, is approximated by β_0 , where β_0 is obtained by setting $s = 0$ in the above equation. Our estimate of β_0 , say $\hat{\beta}_0$ is obtained by performing a least squares analysis using the observations y_{t-m}, \dots, y_{t+m} provided that $2m + 1 \geq p$. If we go through the detailed, though straightforward, steps of this analysis

[Anderson (1971), p. 49], then we see that $\hat{\beta}_0$ is of the general form $\sum_{s=-m}^m c_s y_{t+s}$, with $c_s = c_{-s}$ and $\sum_{s=-m}^m c_s = 1$. The coefficients c_s are polynomials in s depending on m and k , where $k = [p/2]$.^{2/}

For example, if we choose $p = 0$ or 1 , that is, if the polynomial is of degree 0 or 1 , then $\hat{\beta}_0 = \sum_{s=-m}^m y_{t+s} / (2m + 1)$; thus $c_s = (2m + 1)^{-1}$, and we obtain a moving average with equal weights. When $k = 1$, that is, when $p = 2$ or 3 ,

$$\hat{\beta}_0 = \sum_{s=-m}^m \frac{[3(3m^2 + 3m - 1) - 15s^2] y_{t+s}}{(2m - 1)(2m + 1)(2m + 3)}.$$

In Table 3.3 given below, we list the values of c_s 's for $k = 1$, and $m = 2, 3, 4$ and 5 .

Since $\hat{\beta}_0$ is obtained via a least squares analysis, we should bear in mind that if $m < k$, $\hat{\beta}_0$ is undetermined, whereas if $m = k$, we are fitting a $2m + 1$ degree polynomial to $2m + 1$ points and thus $\hat{\beta}_0 = y_t$; that is, the fit is perfect. If $m > k$, the moving average is nontrivial, in the sense that it involves several values of y_{t+s} .

^{2/} [a] denotes the largest integer less than or equal to a .

Table 3.3

Coefficients in Smoothing Formulas for $k = 1$

m	$c_{-5} = c_5$	$c_{-4} = c_4$	$c_{-3} = c_3$	$c_{-2} = c_2$	$c_{-1} = c_1$	c_0
2				$-\frac{3}{35}$	$\frac{12}{35}$	$\frac{17}{35}$
3			$-\frac{2}{21}$	$\frac{3}{21}$	$\frac{6}{21}$	$\frac{7}{21}$
4		$-\frac{21}{231}$	$\frac{14}{231}$	$\frac{39}{231}$	$\frac{54}{231}$	$\frac{59}{231}$
5	$-\frac{36}{429}$	$\frac{9}{429}$	$\frac{44}{429}$	$\frac{69}{429}$	$\frac{84}{429}$	$\frac{89}{429}$

3.2.2 Some Remarks on Smoothing

The main purpose of smoothing is to make the variance of the smoothed sequence y_t^* small relative to the variance of the original sequence y_t . We have seen before that when $k = 1$ ($p = 0$ or 1), $c_s = (2m + 1)^{-1}$, and the variance of y_t^{*2} equals $\sigma^2/(2m + 1)$. In general, it can be shown [Anderson (1971), Theorem 3.3.1] that the variance of y_t^{*2} equals $\sigma^2 c_0$. Since the smoothed value y_t^* can be used as an estimate of the trend at time t , we can also say that the main purpose of smoothing is to estimate the trend of y_t with minimum error. The error consists of two parts:

- i) the bias $E(y_t - y_t^*) = f(t) - \sum_{s=-m}^{s=m} c_s f(t + s)$, and
- ii) the random part $u_t^* = \sum_{s=-m}^{s=m} c_s u_{t+s}$.

For a given value of p , the bias goes up (in most cases) and the variance of the random part goes down as we increase m . For a given value of m , the bias goes down but the variance increases as we increase p . We are therefore faced with the deciding between the values of m and p . One possibility may be to choose those values of m and p that minimize the mean squared error, which is the sum of the variance and the squared bias. However, since σ^2 is unknown, it is difficult to formulate this problem mathematically and give a satisfactory solution. Thus, our choice of m and p must be made on the basis of general experience and intuition.

Another difficulty with smoothing is that to obtain y_t^* , the estimated trend at time t , we have to use y_{t-m}, \dots, y_{t+m} ; thus the first smoothed value is y_{m+1}^* and the last y_{T-m}^* . We therefore do not have an estimate of the trend at the beginning and at the end of the time period.

3.2.3 Examples Illustrating the Practical Value of Smoothing

Example 1

Bhattacharya and Klotz (1966) use freezing dates and thawing dates of Lake Mendota to test for a warming trend. Their data covers a period of 111 years, namely 1854 to 1965. The number of days to freezing in each winter season is measured from Nov. 23. The winter of the season is $t = 1, \dots, 111$ with $t = 1$ denoting 1854-1855. In the interest of keeping our exposition simple, only a portion of the data is considered

by us. The abstracted data is given in Table 3.4. The number of days to freezing measured from Nov. 23, 1853 + t, $t = 1, \dots, 12$, is denoted by y_t . A tendency for y_t to increase with t is interpreted as a warming trend.

In Figure 3.2.1, we show a graph of y_t versus t ; this graph reveals a large amount of variability which tends to conceal a trend. We shall soon see that a graph of the smoothed data (Figure 3.3) reveals the trend more conspicuously.

For illustrative purposes, let us choose $m = 1$ and $p = 1$; thus, we will be fitting a polynomial of degree 1 to 3 successive observations. Note that the cyclical nature of the data suggests that we consider small values of m , unless our goal is to overcome the cycles and just look for a linear trend. For these values of m and p , our smoothed sequence is

$$\begin{aligned} y_t^* &= \sum_{s=-m}^{s=m} c_s y_{t+s} \quad , \quad t = m + 1, \dots, T - m \quad , \\ &= \frac{1}{2m + 1} \sum_{s=-m}^{s=m} y_{t+s} \quad , \quad t = m + 1, \dots, T - m \quad , \\ &= \frac{1}{3} \sum_{s=-1}^{s=1} y_{t+s} \quad , \quad t = 2, \dots, 11 \quad . \end{aligned}$$

In Table 3.4, we show the smoothed values y_t^* obtained by using the above formula.

In Figure 3.3, we show a plot of y_t and y_t^* versus t . Note that the plot of y_t^* shows less fluctuations than the plot of y_t ; this

Table 3.4

Observed and Smoothed Values of the Number of Days to
Freezing of Lake Mendota
(from Nov. 23) 1854, ..., 1865

Time Period t 1853 + t	Number of Days to Freezing, y_t (from Nov. 23)	Smoothed Values y_t^* $m=1, p=1$	Smoothed Values $m=2, p=1$
1	25		
2	13	13.3	
3	2	10.0	13.8
4	15	10.3	13.0
5	14	16.7	12.2
6	21	14.7	18.4
7	9	21.0	20.4
8	33	22.3	20.6
9	25	24.3	20.6
10	15	20.3	23.8
11	21	20.3	
12	25		

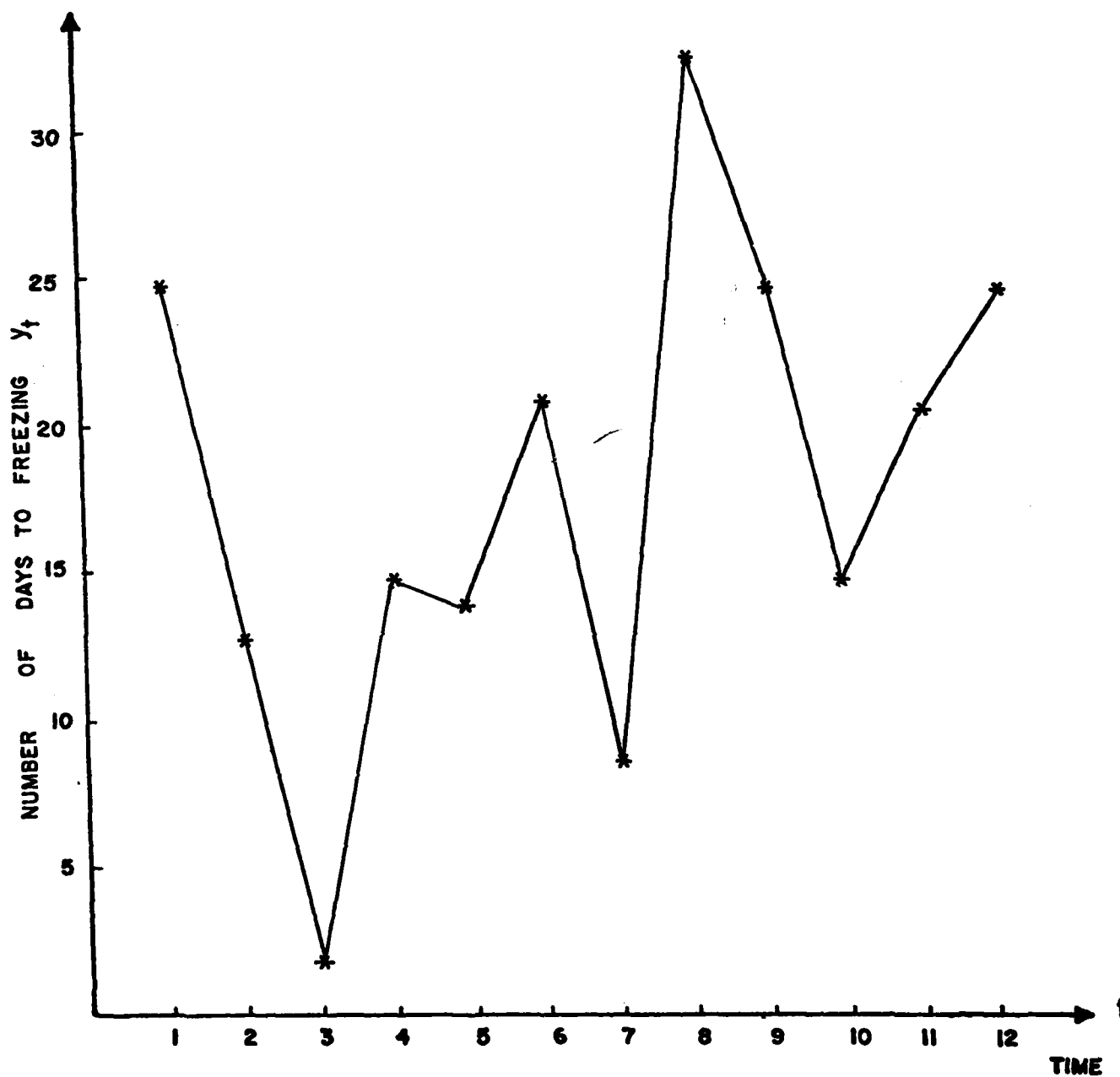


Figure 3.2.1. Number of days to freezing of Lake Mendota against time.

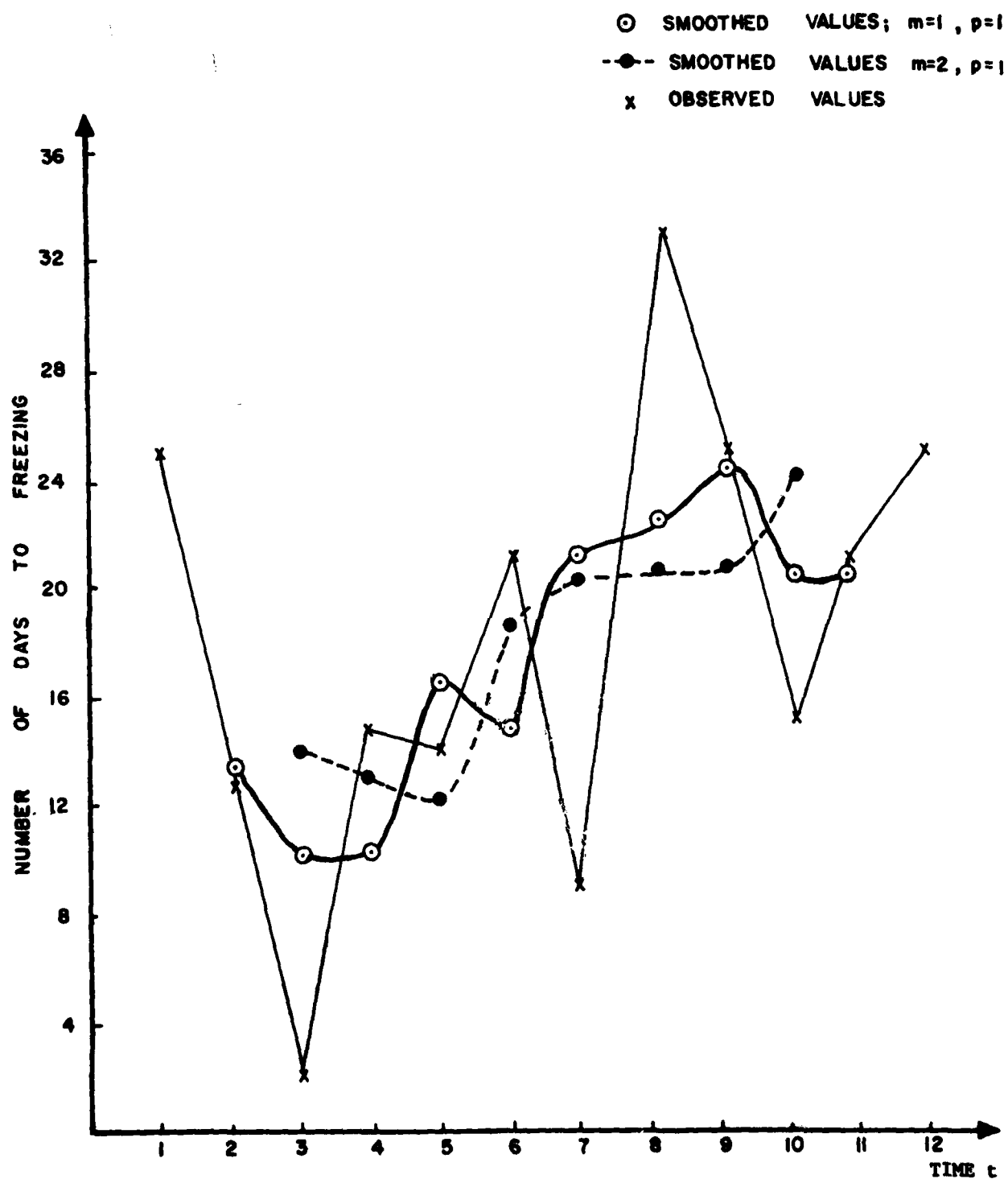


Figure 3.3. Observed and smoothed values of the number of days to freezing of Lake Mendota.

is because the effect of smoothing is to reduce the variability of the y_t . The main advantage of smoothing this data is brought about by the fact that the y_t^* 's reveal an upward trend.

An examination of the entries in Table 3.1 shows that the smoothing is accomplished by pulling closer the values of y_t . For example, the values y_2^* , y_3^* , and y_4^* are closer together than the values of y_2 , y_3 and y_4 , the value $y_3 = 2$ being changed by smoothing to $y_3^* = 10$.

We could continue to smooth this data further by choosing other values of m and p . For example, if we choose $m = 2$ and $p = 1$, then the smoothed values y_t^{**} will be brought still closer; these are shown in column 4 of Table 3.1.

Example 2

We shall now reconsider the data of Table 3.1 on y_t , the annual meat consumption in the U.S. in year t ; this data is graphed in Figure 3.1. Recall that a polynomial in t of degree 3 provides us with a good description of the cyclical pattern in this data. We shall now illustrate how the variability in this data can be reduced by smoothing, with the result that the cyclical pattern becomes more conspicuous. We choose $m = 2$, and $p = 1$ in our smoothing formula, and show the smoothed values y_t^* in column 3 of Table 3.5; also shown in column 2 of Table 3.5 are the actual values y_t . In Figure 3.3.1 we plot the actual and the smoothed values versus t . It is of interest to compare the plot of y_t^* versus t , and the plot of the third degree polynomial fit \hat{y}_t versus t given in Figure 3.2. It appears that y_t^* is closer to \hat{y}_t than the original values y_t .

Table 3.5

Observed and Smoothed Values of the Annual Consumption of
Meat in the United States from 1919-1941

Time Period t 1918 + t	Annual Consumption of Meat y_t	Smoothed Values y_t^* m=2, p=1
1	171.5	-
2	167.0	-
3	164.5	170.34
4	169.3	171.88
5	179.4	173.00
6	179.2	174.20
7	172.6	174.06
8	170.5	171.12
9	168.6	167.88
10	164.7	165.78
11	163.0	163.72
12	162.1	162.24
13	160.2	162.46
14	161.2	162.56
15	165.8	159.48
16	163.5	159.48
17	146.7	158.6
18	160.2	156.8
19	156.8	157.18
20	156.8	162.78
21	165.4	166.48
22	174.7	-
23	178.7	-

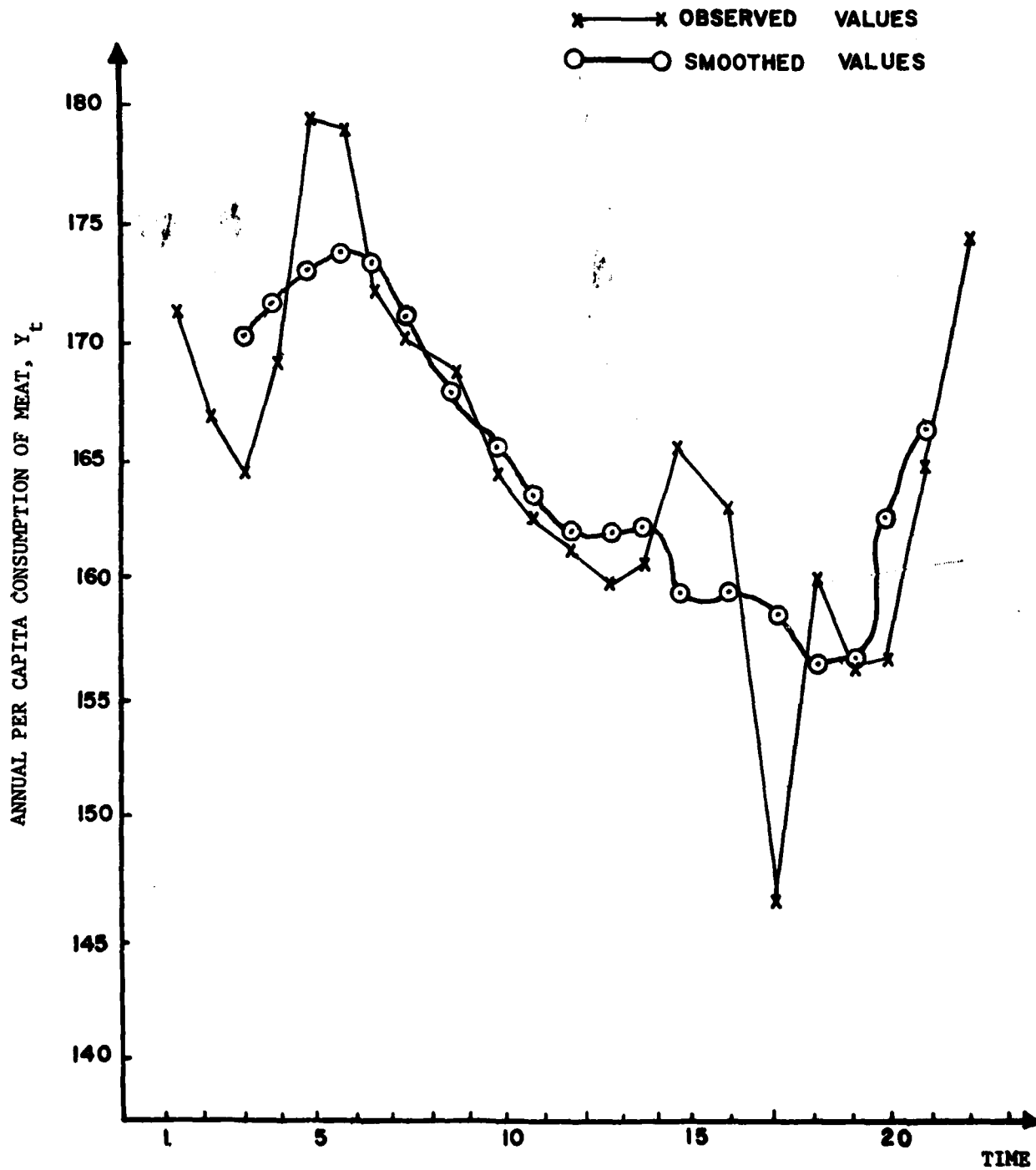


Figure 3.3.1. Observed and smoothed values of the annual consumption of meat in the United States from 1919-1941.

3.2.4 Smoothing and Seasonal Variation

Many economic time series have a seasonal factor, then we want to use the model $y_t = f(t) = g(t) + h(t)$, where $h(t)$ is a trend and $g(t)$ is a periodic function of period n . For example, $n = 4$ when we are dealing with quarterly data such as dividends of A.T.&T. stock, $n = 12$ when we are dealing with monthly data such as the consumer price index, and $n = 52$ when we are dealing with weekly data, such as the yield of treasury bills.

The defining characteristic of the periodic function $g(t)$ is that

$$g(t + n) = g(t) \quad , \quad t = 1, 2, \dots, T - n \quad .$$

We can always normalize the periodic function $g(t)$ so that $\sum_{t=1}^n g(t) = 0$; that is we can center the periodic function about 0. When this is done, we note that $\sum_{t=1}^n g(t + s) = 0$, $s = 0, 1, \dots, T - n$. It is common to choose T to be a multiple of n , say $T = kn$, where k is some integer.

In Figure 3.4(a) we illustrate the behavior of a periodic function centered at some constant C . In Figure 3.4(b) we illustrate the behavior of another periodic function centered at 0.

Suppose that we consider a moving average of n terms (where n is the period) with equal coefficients; that is, we consider

$\frac{1}{n} \sum_{s=1}^n y_{t+s}$. Recall that the weights $c_s = \frac{1}{n}$ when we smooth over n observations and choose $p = 0$ or 1 ($m = (n - 1)/2$). Then

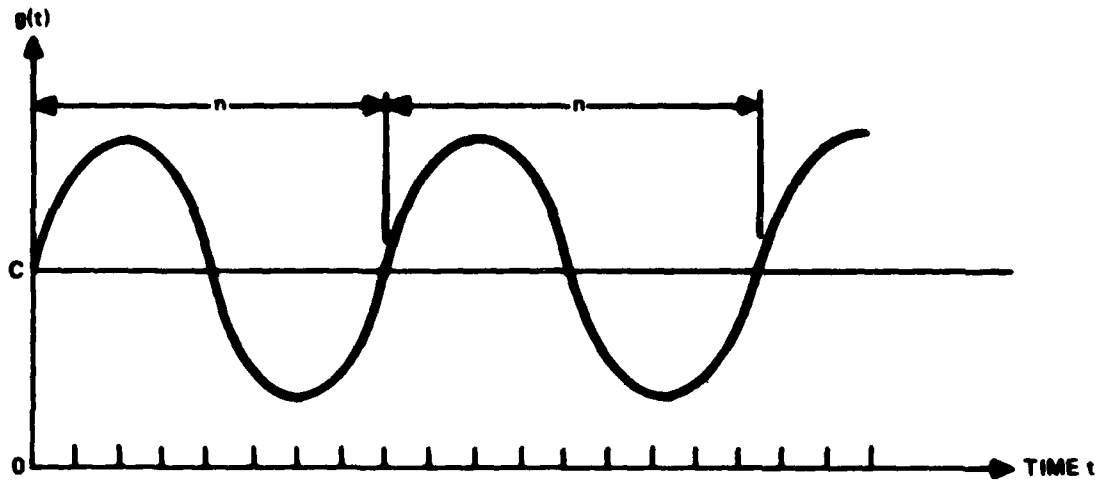


FIGURE 3.4A SINUSOIDAL PERIODIC FUNCTION $g(t)$ CENTERED AT C WITH PERIOD n

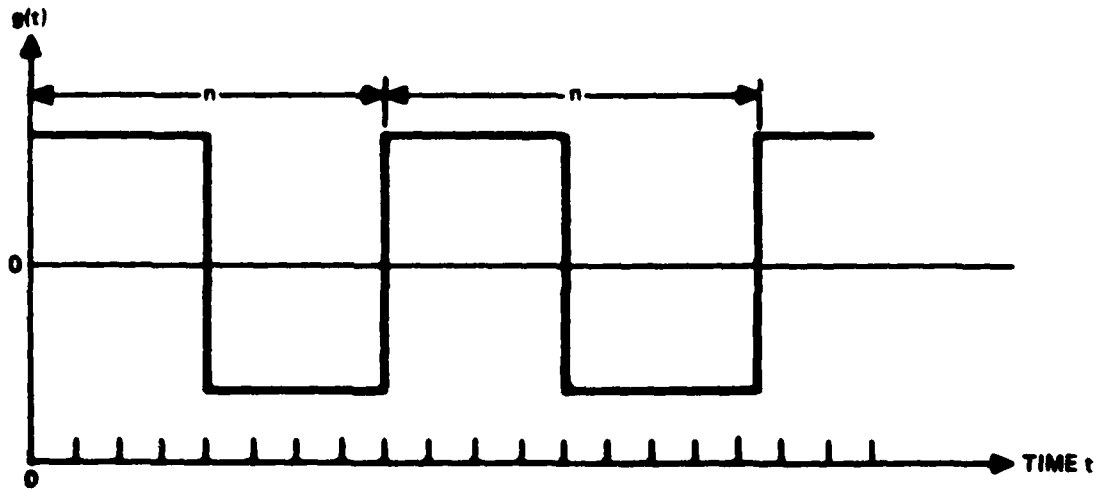


FIGURE 3.4B RECTANGULAR PERIODIC FUNCTION $g(t)$ CENTERED AT 0 WITH PERIOD n

$$\mathbb{E} \frac{1}{n} \sum_{s=1}^n y_{t+s} = \frac{1}{n} \sum_{s=1}^n (h(t+s) + g(t+s)) = \frac{1}{n} \sum_{s=1}^n h(t+s)$$

since a centering of the periodic function $g(t)$ ensures that

$$\frac{1}{n} \sum_{s=1}^n g(t+s) = 0.$$

Thus a moving average of n terms with equal weights will eliminate the seasonal variation $g(t)$, whenever the period of $g(t)$ is n . This procedure is sometimes proposed to eliminate the effect of a cyclical movement when estimating the trend.

If n is even, that is, if $n = 2m$, we use

$$y_t^* = \frac{1}{2m} \left[\sum_{s=-(m-1)}^{s=(m-1)} y_{t+s} + \frac{1}{2} y_{t-m} + \frac{1}{2} y_{t+m} \right], \quad t = m+1, \dots, T-m.$$

Then

$$\begin{aligned} \mathbb{E} y_t^* &= \frac{1}{2m} \left[\sum_{s=-(m-1)}^{m-1} f(t+s) + \frac{1}{2} f(t-m) + \frac{1}{2} f(t+m) \right] \\ &= \frac{1}{2m} \left[\sum_{s=-(m-1)}^{m-1} h(t+s) + \frac{1}{2} h(t-m) + \frac{1}{2} h(t+m) \right], \end{aligned}$$

since $(1/4m)g(t-m) = (1/4m)g(t+m)$. If $h(t)$ is changing, slowly, then $\mathbb{E} y_t^*$ will be close to $h(t)$; in particular if $h(t)$ is linear, $\mathbb{E} y_t^*$ will equal $h(t)$.

When $T = kn$, we can define $g(t)$ uniquely by

$$g(t) = \frac{1}{k} \sum_{j=0}^{k-1} f(t + nj) - \frac{1}{T} \sum_{s=1}^T f(s), \quad t = 1, 2, \dots, n.$$

For example, the seasonal effect of December is the difference between the average of all Decembers and the over-all average. Thus, an estimate of $g(t)$ is

$$\frac{1}{k} \sum_{j=0}^{k-1} y_{t+nj} - \frac{1}{T} \sum_{s=1}^T y_s, \quad t = 1, 2, \dots, n.$$

Clearly, this estimator is unbiased. Its variance can be calculated as follows:

$$\begin{aligned} \text{Var} \left[\frac{1}{k} \sum_{j=0}^{k-1} y_{t+nj} - \frac{1}{T} \sum_{s=1}^T y_s \right] &= \text{Var} \left[\left(\frac{1}{k} - \frac{1}{T} \right) \sum_{j=0}^{k-1} y_{t+nj} \right] \\ &\quad + \text{Var} \left[\frac{1}{T} \sum_{s \neq t+nj}^T y_s \right] \end{aligned}$$

since the common terms in the above two summation signs have been absorbed under one summation. Thus

$$\begin{aligned} \text{Var} \left[\frac{1}{k} \sum_{j=0}^{k-1} y_{t+nj} - \frac{1}{T} \sum_{s=1}^T y_s \right] &= \left(\frac{1}{k} - \frac{1}{T} \right)^2 k \sigma^2 + \left(\frac{1}{T} \right)^2 (T - k) \sigma^2 \\ &= \sigma^2 \frac{n-1}{T}. \end{aligned}$$

3.3 The Variate Difference Method

The variate difference method is sometimes used to estimate the variance of the (uncorrelated) error term u_t when the trend $f(t)$ is smooth; that is, when $f(t)$ can be reasonably well approximated by a polynomial of a low degree. Another use of variate differences is to test for the lack of correlation.

In the variate difference method, we consider successive differences of elements in a time series. We shall need to define an operator P where $Pu_t = u_{t+1}$, for $t = \dots, -1, 0, 1, \dots$. Given a sequence $\{u_t\}$, $t = \dots, -1, 0, 1, \dots$, $P\{u_t\} = \{u_{t+1}\}$; thus the operator P when applied to a sequence gives us another sequence in which the subscripts are shifted by 1.

P is said to be a linear operator because

i) for each sequence $\{u_t\}$, and each real number c

$$P\{cu_t\} = cP\{u_t\}, \text{ and}$$

ii) for each pair of sequences $\{u_t\}$ and $\{v_t\}$

$$P\{u_t + v_t\} = P\{u_t\} + P\{v_t\}.$$

We shall define $P^0 u_t = u_t$ and $P^n(u_t) = P(P^{n-1}u_t)$, $n = 2, 3, \dots$; thus, we can verify that $P^n(u_t) = u_{t+n}$. Also, by definition $(cP)u_t = cP(u_t)$, and $(c_1P^{n_1} + c_2P^{n_2} + \dots + c_kP^{n_k})u_t = c_1P^{n_1}u_t + c_2P^{n_2}u_t + \dots + c_kP^{n_k}u_t = c_1u_{t+n_1} + c_2u_{t+n_2} + \dots + c_ku_{t+n_k}$.

Consider the forward difference operator Δ , where $\Delta u_t = u_{t+1} - u_t$; since $\Delta u_t = u_{t+1} - u_t = Pu_t - u_t = (P-1)u_t$, it follows that $(P-1)u_t = \Delta u_t$ or that $\Delta = P - 1$. The second-order forward difference operator $\Delta^2 u_t$ is $\Delta^2 u_t = \Delta(\Delta u_t) = \Delta(u_{t+1} - u_t) = \Delta u_{t+1} - \Delta u_t = u_{t+2} - 2u_{t+1} + u_t$. Equivalently, $\Delta^2 u_t = (P-1)^2 u_t = (P^2 - 2P + 1)u_t = u_{t+2} - 2u_{t+1} + u_t$, where in squaring $(P-1)$ we treat P as a real coefficient. In general, we have $\Delta^r u_t = (P-1)^r u_t = \sum_{j=0}^r (-1)^{r-j} \binom{r}{j} P^j u_t = \sum_{j=0}^r (-1)^{r-j} \binom{r}{j} u_{t+j}$.

Consider a polynomial in t of degree p ; for example, let $f(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_p t^p$, and take its first forward difference

$$\begin{aligned}\Delta f(t) &= \Delta(\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_p t^p) \\ &= \alpha_0 + \alpha_1(t+1) + \alpha_2(t+1)^2 + \dots + \alpha_p(t+1)^p - \alpha_0 \\ &\quad - \alpha_1 t - \alpha_2 t^2 - \dots - \alpha_p t^p \\ &= \alpha_1 + \alpha_2((t+1)^2 - t^2) + \dots + \alpha_p((t+1)^p - t^p) \\ &= p\alpha_p t^{p-1} + (\text{constant})t^{p-2} + \dots + (\alpha_p + \alpha_{p-1} + \dots + \alpha_1) .\end{aligned}$$

Thus, by taking the first forward difference, we have reduced the degree of the polynomial by 1. In general, if $f(t)$ is a polynomial in degree p , then

$$\Delta^r f(t) = 0, \quad r = p+1, p+2, \dots$$

Thus, if we consider a trend to be approximated by a polynomial of degree p , then by taking $p+1$ or more forward differences of the trend, we can reduce the trend to 0, approximately.

3.3.1 Taking Differences of the Observed Series

Suppose that we have an observed series $\{y_t\}$ which will be treated as composed of a trend $f(t)$ and a random error u_t . Since $\Delta = (P - 1)$, Δ is a linear operator,

$$\Delta y_t = \Delta(f(t) + u_t) = \Delta f(t) + \Delta u_t .$$

If $f(t)$ is a polynomial in t of degree p , where $p < r$, then $\Delta^r f(t) = 0$; thus $\Delta^r y_t = \Delta^r f(t) + \Delta^r u_t = \Delta^r u_t$. It now follows that

$$\Delta^r y_t = 0 .$$

This is important because it shows one method of eliminating a (polynomial) trend.

The differencing also affects the variance; we have

$$\text{Var} (\Delta^r y_t) = \text{Var} (\Delta^r u_t) = \text{Var} ((P - 1)^r u_t) .$$

In order to obtain $\text{Var} ((P - 1)^r u_t)$, we note that

$$\begin{aligned} (P - 1)^r u_t &= [P^r - \binom{r}{1} P^{r-1} + \dots + (-1)^r] u_t \\ &= u_{t+r} - \binom{r}{1} u_{t+r-1} + \dots + (-1)^r u_t . \end{aligned}$$

Thus $\text{Var} (\Delta^r y_t) = \sigma^2 [1 + \binom{r}{1}^2 + \dots + (-1)^{2r}] = \sigma^2 \binom{2r}{r}$. For a proof of the last equality given above, we refer the reader to Anderson (1971), p.64.

The above results enable us to propose an estimator of σ^2 , say v_r , where

$$v_r = \frac{\sum_{t=1}^{T-r} (\Delta^r y_t)^2}{(T - r) \binom{2r}{r}} .$$

4. Cyclical Trends

In many time series, especially those involving economic data, the trend $f(t)$ is a periodic function of time; that is $f(t + \phi) = f(t)$, for some period ϕ . Thus, given the function $f(t)$ on any interval of length ϕ , we can determine it over its entire range. We shall consider here the analysis of a time series when the trend $f(t)$ can be specified in terms of linear combinations of sines and cosines; this seems like a natural way to consider periodic trends. However, we shall first need the following preliminary notions.

4.1 Transformations and Representations

In this section we shall see some alternate ways of representing any sequence of T numbers y_1, y_2, \dots, y_T (not necessarily a time series), and a periodic function $f(t)$. In Section 4.2 we shall return to the analysis of time series by applying some of the techniques discussed in this section.

4.1.1 Trigonometric Functions and their Orthogonality

The trigonometric functions $\sin t$ and $\cos t$ are periodic with period 2π ; that is

$$\sin(t + 2\pi) = \sin t, \quad \cos(t + 2\pi) = \cos t.$$

Furthermore, for any k , $k = 0, \pm 1, \pm 2, \dots$,

$$\sin(t + 2\pi k) = \sin t, \quad \cos(t + 2\pi k) = \cos t.$$

Suppose that we wish to make a linear transformation of the argument t in $\sin(\cdot)$ and $\cos(\cdot)$ by multiplying it by some constant λ . Then, $\sin \lambda t$ and $\cos \lambda t$ will also be periodic, but the period is now $2\pi/\lambda$. That is,

$$\sin\left(\lambda\left(t + \frac{2\pi}{\lambda}\right)\right) = \sin \lambda t \quad , \quad \cos\left(\lambda\left(t + \frac{2\pi}{\lambda}\right)\right) = \cos \lambda t \quad .$$

The effect of λ is to expand or to contract the time scale; small values of λ expand the time scale whereas large values of λ contract it.

(See Figure 4.1.)

The reciprocal of the period is called the frequency. The frequency need not be integer valued. The frequency denotes the number of periods in a unit interval.

In dealing with the periodic functions $\sin t$ and $\cos t$, we may want to shift (or translate) the entire sine or the cosine curve. This is accomplished by the introduction of a shift parameter θ . More specifically,

$$\sin\left(\lambda\left(t + \frac{2\pi}{\lambda}\right) - \theta\right) = \sin(\lambda t + 2\pi - \theta) = \sin(\lambda t - \theta) \quad , \quad \text{and}$$

$$\cos\left(\lambda\left(t + \frac{2\pi}{\lambda}\right) - \theta\right) = \cos(\lambda t + 2\pi - \theta) = \cos(\lambda t - \theta) \quad .$$

Note that since the maximum of $\cos \lambda t$ occurs at $\lambda t = 2\pi k$, $k = 0, \pm 1, \pm 2, \dots$, the maximum of $\cos(\lambda t - \theta)$ occurs at $\lambda t = 2\pi k + \theta$; that is, at $t = (\theta + 2\pi k)/\lambda$. The angle θ is called the phase. Usually we choose θ in such a manner that the first maximum occurs at $t = \theta/\lambda$. At $t = 0$ the function is either $\cos \theta$ or $-\sin \theta$.

In Figure 4.1 we illustrate the above properties via a cosine function with a frequency of $1/2\pi$, $1/(4\pi)$, and $1/\pi$ by suitable choices of λ . The effect of the phase is indicated by the dotted lines; we have chosen θ , θ , and $\pi/2$ for the phases.

Since $\cos(a - b) = \cos a \cos b + \sin a \sin b$, we have, for any constant ρ ,

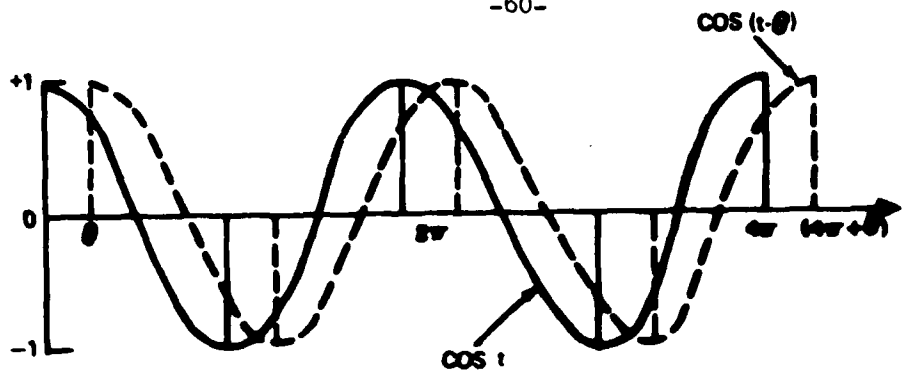
$$\begin{aligned}\rho(\cos(\lambda t - \theta)) &= \rho(\cos \lambda t \cos \theta + \sin \lambda t \sin \theta) \\ &= \alpha \cos \lambda t + \beta \sin \lambda t,\end{aligned}$$

where $\alpha = \rho \cos \theta$ and $\beta = \rho \sin \theta$. Since $\cos^2 \theta + \sin^2 \theta = 1$, $\rho^2 = \alpha^2 + \beta^2$, and since $\tan \theta = \sin \theta / \cos \theta$, $\theta = \tan^{-1}(\beta/\alpha)$. The maximum value of the function $\rho \cos(\lambda t - \theta)$ is ρ ; ρ is therefore called the amplitude of the function. The quantity ρ^2 is called the intensity.

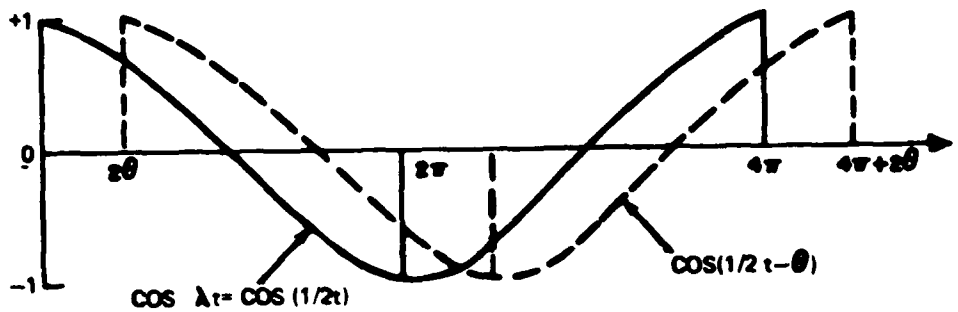
In the light of the above discussion and Figure 4.1 we remark that by a suitable choice of λ , θ , and ρ , we can obtain any desired shape of the cosine curve. The same is also true of the sine curve. We shall make use of this geometric property of the trigonometric functions in Section 4.1.3, wherein we approximate a periodic function $f(t)$ by an infinite linear combination of sines and cosines with varying amplitudes and frequencies.

The Orthogonality of Trigonometric Functions

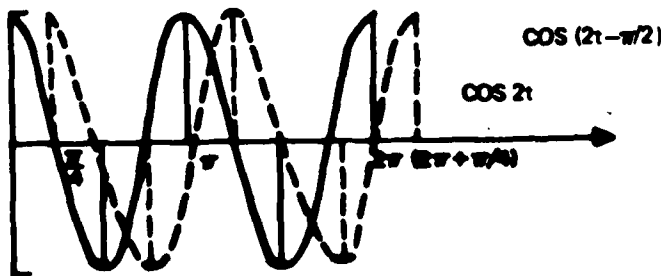
An advantage of the trigonometric functions is that they exhibit a certain type of an orthogonality property. This property makes it



COSINE FUNCTIONS WITH PERIOD 2π , FREQUENCY $1/2\pi$ AND PHASE 0 , OR θ
(THE MAXIMA OCCUR AT $t=0, 2\pi, 4\pi, \dots$, OR AT $\theta, 2\pi + \theta, 4\pi + \theta, \dots$.)



COSINE FUNCTION WITH PERIOD 4π ($\lambda = 1/2$), FREQUENCY $1/(4\pi)$, AND PHASE 0 OR θ
(THE MAXIMA OCCUR AT $t=0, 4\pi, 8\pi, \dots$, OR AT $2\theta, 4\pi + 2\theta, \dots$.)



COSINE FUNCTION WITH PERIOD π ($\lambda = 2$), FREQUENCY $1/\pi$ AND PHASE 0 AND $\pi/2$
(MAXIMA OCCUR AT $\lambda t = 0, 2\pi, 4\pi, \dots$, OR AT $\pi/4, 2\pi + \pi/4, \dots$.)

FIGURE 4.1 COSINE CURVE WITH VARYING FREQUENCIES AND PHASE

convenient for us to work with them. We shall merely state this property here and refer the reader to Anderson (1971), p. 94, for a proof of the pertinent results.

Consider a series of length T , and let

$$\begin{aligned} \left[\frac{1}{2}T\right] &= \frac{T}{2} \quad , \quad \text{if } T \text{ is even} \\ &= \frac{T-1}{2} \quad , \quad \text{if } T \text{ is odd} \quad . \end{aligned}$$

Then, the orthogonality property specifies that

$$(4.1) \quad \sum_{t=1}^T \cos \frac{2\pi j}{T} t \cos \frac{2\pi k}{T} t = \begin{cases} 0 & , \quad 0 \leq k \neq j \leq \left[\frac{1}{2}T\right] \quad , \\ \frac{1}{2}T & , \quad 0 < k = j < \frac{1}{2}T \quad , \\ T & , \quad k = j = 0, \text{ or } \frac{1}{2}T \quad . \end{cases}$$

$$(4.2) \quad \sum_{t=1}^T \cos \frac{2\pi j}{T} t \sin \frac{2\pi k}{T} t = 0 \quad , \quad k, j = 0, 1, \dots, \left[\frac{1}{2}T\right] \quad ,$$

$$(4.3) \quad \sum_{t=1}^T \sin \frac{2\pi j}{T} t \sin \frac{2\pi k}{T} t = \begin{cases} 0 & , \quad 0 \leq k \neq j \leq \left[\frac{1}{2}T\right] \quad , \\ \frac{1}{2}T & , \quad 0 < k = j < \frac{1}{2}T \quad , \\ 0 & , \quad k = j = 0, \text{ or } \frac{1}{2}T \quad . \end{cases}$$

In the above expressions, we are considering T sums of cosine and sine functions of the form $\cos \lambda t$ and $\sin \lambda t$, where λ is to be identified with $2\pi j/T$. Since the frequency of $\cos \lambda t$ and $\sin \lambda t$ is $\lambda/2\pi$, the appropriate frequencies in the above equations are j/T , $j = 0, 1, \dots, \left[\frac{1}{2}T\right]$, and their periods are T/j . When (4.1) through (4.3)

hold, we say that the T cosine and sine functions with frequencies j/T are orthogonal to each other.

4.1.2 The Fourier Representation of any Finite Sequence of Numbers

Consider any sequence of T numbers y_1, y_2, \dots, y_T , where T is even. These numbers need not be the observed values of a time series. The T numbers define the coordinates of a point in a space of T dimensions. We would like to refer to this point in another coordinate system. We shall do this by using the orthogonal trigonometric functions discussed above.

Motivated by the result of (4.1)-(4.3), we shall define a $T \times T$ matrix M for T even by

$$M = \sqrt{\frac{2}{T}} \begin{vmatrix} \frac{1}{\sqrt{2}} & \cos \frac{2\pi}{T} & \sin \frac{2\pi}{T} & \cos \frac{4\pi}{T} & \dots & \sin \frac{2\pi}{T}(\frac{1}{2}T-1) & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \cos \frac{2\pi}{T}2 & \sin \frac{2\pi}{T}2 & \cos \frac{4\pi}{T}2 & \dots & \sin \frac{4\pi}{T}(\frac{1}{2}T-1) & \frac{1}{\sqrt{2}} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \frac{1}{\sqrt{2}} & 1 & 0 & 1 & & 0 & \frac{1}{\sqrt{2}} \end{vmatrix}$$

From the orthogonality relationships, we have $M'M = I$. Let

$\underline{y} = (y_1, y_2, \dots, y_T)'$ and $\underline{x} = (x_1, x_2, \dots, x_T)'$, where $\underline{x} = M'\underline{y}$. Since $\underline{M}\underline{M}' = I$, we have $\underline{y} = \underline{M}\underline{x}$, where $\underline{x} = M'\underline{y}$ gives us

$$x_1 = \frac{1}{\sqrt{T}} \sum_{t=1}^T y_t ,$$

$$x_{2k} = \sqrt{\frac{2}{T}} \sum_{t=1}^T y_t \cos \frac{2\pi k}{T} t, \quad k = 1, 2, \dots, \frac{1}{2}T - 1$$

$$x_{2k+1} = \sqrt{\frac{2}{T}} \sum_{t=1}^T y_t \sin \frac{2\pi k}{T} t, \quad k = 1, 2, \dots, \frac{1}{2}T - 1,$$

$$x_T = \frac{1}{\sqrt{T}} \sum_{t=1}^T y_t (-1)^t.$$

Thus, using $y = Mx$ we can write $y_t, t = 1, 2, \dots, T$ as

$$(4.4) \quad y_t = \sqrt{\frac{2}{T}} \left(\frac{1}{\sqrt{2}} x_1 + x_2 \cos \frac{2\pi}{T} t + \dots + x_T \frac{(-1)^t}{\sqrt{2}} \right).$$

Equation (4.4) is known as the Fourier representation of y_1, y_2, \dots, y_T with discrete Fourier coefficients, x_1, x_2, \dots, x_T .

When T is odd, we go through an analogous development except

that the last column of M is $\left(\sin \left[\frac{2\pi \frac{1}{2}(T-1)}{T} \right], \sin \left[\frac{2\pi \frac{1}{2}(T-1)}{T} \right], \dots, 0 \right)$,
and now

$$(4.5) \quad y_t = \sqrt{\frac{2}{T}} \left(\frac{1}{\sqrt{2}} x_1 + x_2 \cos \frac{2\pi}{T} t + \dots + x_T \sin \frac{2\pi \frac{1}{2}(T-1)}{T} t \right),$$

$$t = 1, 2, \dots, T.$$

Periodic Sequences

In time series analysis, the sequence of numbers y_1, y_2, \dots, y_T may be periodic with a period n . Thus

$$y_{t+n} = y_t, \quad t = 1, 2, \dots, T - n,$$

for some integer n . Suppose that T is a multiple of n , say $T = hn$. We can represent this sequence in terms of only n trigonometric functions as follows:

When n is even, then for $t = 1, 2, \dots, n$

$$(4.6) \quad y_t = \sqrt{\frac{2}{n}} \left(\frac{1}{\sqrt{2}} x_1^* + x_2^* \cos \frac{2\pi}{n} t + \dots + x_n^* \frac{(-1)^t}{\sqrt{2}} \right),$$

where

$$x_1^* = \frac{1}{\sqrt{n}} \sum_{t=1}^n y_t,$$

$$x_{2k}^* = \sqrt{\frac{2}{n}} \sum_{t=1}^n y_t \cos \frac{2\pi k}{n} t, \quad k = 1, \dots, \left[\frac{1}{2}(n-1) \right],$$

$$x_{2k+1}^* = \sqrt{\frac{2}{n}} \sum_{t=1}^n y_t \sin \frac{2\pi k}{n} t, \quad k = 1, \dots, \left[\frac{1}{2}(n-1) \right],$$

$$x_n^* = \frac{1}{\sqrt{n}} \sum_{t=1}^n y_t (-1)^t.$$

Since the cosine and the sine functions are periodic

$$\begin{aligned} y_{t+n} &= y_t = \sqrt{\frac{2}{n}} \left(\frac{1}{\sqrt{2}} x_1^* + x_2^* \cos \frac{2\pi}{n} t + \dots + x_n^* \frac{(-1)^t}{\sqrt{2}} \right) \\ &= \sqrt{\frac{2}{n}} \left(\frac{1}{\sqrt{2}} x_1^* + x_2^* \cos \frac{2\pi}{n} (t+n) + \dots + x_n^* \frac{(-1)^{t+n}}{\sqrt{2}} \right). \end{aligned}$$

Thus (4.6) holds for every t , $t = 1, 2, \dots, T$. When n is odd, the term involving $(-1)^t$ is omitted.

4.1.3 Fourier Representation of a Periodic Function

Consider any periodic function $f(t)$ whose period is ϕ ; let $f(t)$ be defined for all values of t . Thus

$$f(t) = f(t + \phi) = f(t + 2\phi) = \dots$$

We would like to represent $f(t)$ in terms of sine and cosine functions whose period is also ϕ . That is, we would like to write $f(t)$ in terms of an infinite linear combination of $\cos \frac{2\pi}{\phi} 0t (= 1)$, $\sin \frac{2\pi}{\phi} 0t (= 0)$, $\cos \frac{2\pi}{\phi} (t)$, $\sin \frac{2\pi}{\phi} (t)$, $\cos \frac{2\pi}{\phi} (2t)$, $\sin \frac{2\pi}{\phi} (2t)$, For some constants $\alpha_i, \beta_i, i = 0, 1, 2, \dots$, let us consider an infinite series of the form

$$\alpha_0 + \alpha_1 \cos \frac{2\pi}{\phi} t + \beta_1 \sin \frac{2\pi}{\phi} t + \alpha_2 \cos \frac{4\pi}{\phi} t + \beta_2 \sin \frac{4\pi}{\phi} t + \dots$$

Our motivation for considering this infinite sum should be apparent from an examination of Figure 4.2. We show there three sine curves with periods $\phi, \phi/2$, and $\phi/4$, and amplitudes $\beta_1 = 1, \beta_2 = 3/4$ and $\beta_3 = 1/2$ respectively. The sum $\sum_{i=1}^3 \beta_i \sin \frac{2\pi}{\phi} (it)$, shown by the dotted lines of Figure 4.2, could be considered as an approximation to some $f(t)$. However, by a suitable choice of $\beta_i, i = 1, 2, \dots$, the infinite sum $\sum_{i=1}^{\infty} \beta_i \sin \frac{2\pi}{\phi} (it)$ would be a better approximation to $f(t)$. A similar type of an argument leads us to consider an infinite sum of the form $\sum_{i=0}^{\infty} \alpha_i \cos \frac{2\pi}{\phi} (it)$.

In general, we can consider an infinite sum of the pairs of sine and cosine terms

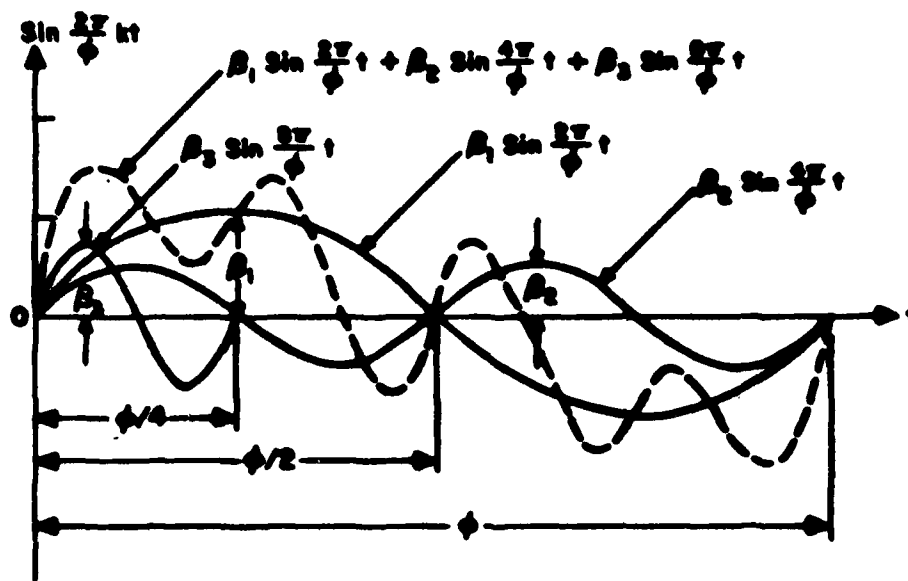


FIGURE 4.2. ILLUSTRATION OF 3 SINE CURVES WITH FREQUENCIES ϕ , $\phi/2$, AND $\phi/4$, AND AMPLITUDES B_1 , B_2 , AND B_3 (RESPECTIVELY), AND THEIR SUM.

$$\sum_{j=0}^{\infty} (\alpha_j \cos \frac{2\pi}{\phi} jt + \beta_j \sin \frac{2\pi}{\phi} jt)$$

to provide us with a still better approximation to $f(t)$.

If the infinite series converges to $f(t)$ for some value of t , then it also converges to $f(t + \phi)$ since $\cos \frac{2\pi}{\phi} k(t + \phi)$ and $\sin \frac{2\pi}{\phi} k(t + \phi)$ are also periodic functions. Thus the infinite sum of trigonometric functions given above is also periodic with period ϕ . We can verify using the trigonometric identities [see Anderson (1971), p. 100] that the functions in the infinite series have the following properties:

$$\int_0^{\phi} \cos^2 \frac{2\pi j}{\phi} t dt = \int_0^{\phi} \sin^2 \frac{2\pi j}{\phi} t dt = \frac{1}{2} \phi, \quad j \neq 0;$$

$$\int_0^{\phi} \cos \frac{2\pi j}{\phi} t \cos \frac{2\pi k}{\phi} t dt = \int_0^{\phi} \sin \frac{2\pi j}{\phi} t \sin \frac{2\pi k}{\phi} t dt = 0, \quad j \neq k;$$

$$\int_0^{\phi} \cos \frac{2\pi j}{\phi} t \sin \frac{2\pi k}{\phi} t dt = 0, \quad \text{all } j \text{ and } k.$$

Under some very general conditions on $f(t)$, Fourier analysis tells us that the infinite series considered here converges to $f(t)$ at every continuity point of $f(t)$. When this happens, and if term by term integration is permissible, then

$$\begin{aligned} \int_0^{\phi} f(t) \cos \frac{2\pi k}{\phi} t dt &= \alpha_0 \int_0^{\phi} \cos \frac{2\pi k}{\phi} t dt \\ &+ \int_0^{\phi} \sum_{j=1}^{\infty} (\alpha_j \cos \frac{2\pi j}{\phi} t + \beta_j \sin \frac{2\pi j}{\phi} t) \cos \frac{2\pi k}{\phi} t dt \\ &= \frac{1}{2} \phi \alpha_k, \quad \text{for } k \neq 0. \end{aligned}$$

The above equation determines α_k as

$$(4.7) \quad \alpha_k = \frac{2}{\phi} \int_0^{\phi} f(t) \cos \frac{2\pi k}{\phi} t dt, \quad k \neq 0.$$

Also

$$(4.8) \quad \alpha_0 = \frac{1}{\phi} \int_0^{\phi} f(t) dt.$$

In a similar manner, multiplication by $\sin \frac{2\pi k}{\phi} t$ gives us

$$(4.9) \quad \beta_k = \frac{2}{\phi} \int_0^{\phi} f(t) \sin \frac{2\pi k}{\phi} t dt, \quad k \neq 0.$$

When the Fourier coefficients α_k , α_0 , and β_k are chosen according to equations (4.7), (4.8) and (4.9) the series is said to represent $f(t)$.

In time series analysis, the function $f(t)$ is used as a trend function in an error model of the form $y_t = f(t) + u_t$. Thus only the values of $f(t)$ at $t = 1, 2, \dots, T$ are relevant. If $f(t)$ is assumed to be periodic with period n , where n is an integer, then only the n values $f(1), f(2), \dots, f(n)$ appear in our analysis. In such a case the

function can be represented at $t = 1, 2, \dots$, by a linear combination of n trigonometric functions as was done in Section 4.1.2.

Summary of Section 4.1

The main point of the discussion in Section 4.1 is that any finite sequence of observations, or any periodic function $f(t)$, can be alternatively represented by a linear combination of trigonometric functions with different frequencies. Specifically,

- 1) Any finite sequence of T observations y_1, y_2, \dots, y_T , periodic or otherwise, can be transformed into a finite set of Fourier coefficients x_1, x_2, \dots, x_T by the use of the orthogonal matrix M .
- 2) Any periodic function $f(t)$ defined for all values of t can be represented by an infinite sum of Fourier terms whose coefficients α_k and β_k are the trigonometric integrals given by equations (4.7) through (4.9).

The need for these alternate representations will be evident in Section 4.2 wherein we will discuss the statistical estimation of cyclical trends. The coefficients x_1, \dots, x_T , the α_k and the β_k can be given some physical interpretations; these too will be discussed later. However, before we proceed to Section 4.2, we shall first illustrate the methodology of this section via some examples.

4.1.4 Examples Illustrating the Use of Fourier Representations

We shall consider here three examples. The first two examples pertain to observations from a real life time series, and the third one

pertains to an arbitrary function. The first two examples illustrate the methodology of Section 4.1.2 whereas the latter illustrates the methodology of Section 4.1.3.

Example 1

In Table 4.1 we present some data on Wolfer's Sunspot Numbers rounded to the nearest integers from the year 1911 through 1933. These numbers have been taken from Table A.3.1 of Anderson (1971). We would like to obtain a Fourier representation of these numbers denoted by us as y_t , for $t = 1, 2, \dots, 33$. A graph of y_t versus t is shown in Figure 4.2.1.

We will have to obtain 33 Fourier coefficients x_1, x_2, \dots, x_{33} in order to obtain the desired Fourier representation. We shall first compute

$$x_1 = \frac{1}{\sqrt{33}} \sum_{t=1}^{33} y_t = 255.894$$

$$x_{2k} = \sqrt{\frac{2}{33}} \sum_{t=1}^{33} y_t \cos \frac{2\pi k}{T} t, \quad k = 1, 2, \dots, 16,$$

$$x_{2k+1} = \sqrt{\frac{2}{33}} \sum_{t=1}^{33} y_t \sin \frac{2\pi k}{T} t, \quad k = 1, 2, \dots, 16.$$

The computed values of x_1, \dots, x_{33} are given in column 4 of Table 4.1. Since $T (= 33)$ is odd, the Fourier representation of y_t , $t = 1, 2, \dots, 33$ is given as

$$y_t = \sqrt{\frac{2}{33}} \left[\frac{x_1}{\sqrt{2}} + \sum_{k=1}^{16} (x_{2k} \cos \frac{2\pi k}{33} t + x_{2k+1} \sin \frac{2\pi k}{33} t) \right].$$

Table 4.1

Wolfer's Sunspot Numbers and Their Fourier Representation

Time t	Wolfer's Sunspot Numbers in the Year $1910 + t$, y_t	Fourier Representa- tion of y_t	Fourier Coefficients x_k Corresponding to Frequency $k/33$	
			x_k	k
1	6	6.000	255.894	1
2	4	4.000	4.874	2
3	1	1.000	-25.832	3
4	10	10.000	-39.242	4
5	47	47.000	-54.309	5
6	57	57.001	-129.425	6
7	104	104.000	-102.646	7
8	81	80.999	25.872	8
9	64	64.000	27.497	9
10	38	37.999	23.536	10
11	26	26.000	20.414	11
12	14	13.999	15.968	12
13	6	6.000	7.612	13
14	17	17.001	-9.711	14
15	44	44.002	-9.160	15
16	64	64.000	-11.522	16
17	69	69.001	-8.382	17
18	78	78.000	6.500	18
19	65	64.999	8.901	19
20	36	36.000	-4.257	20
21	21	21.000	1.181	21
22	11	11.000	-1.477	22
23	6	6.000	2.558	23
24	9	9.000	-4.521	24
25	36	36.002	-8.807	25
26	80	80.003	3.415	26
27	114	114.001	-6.367	27
28	110	109.998	9.473	28
29	89	88.999	-1.577	29
30	68	67.998	2.295	30
31	48	47.999	10.961	31
32	31	30.998	-7.730	32
33	16	16.000	6.564	33

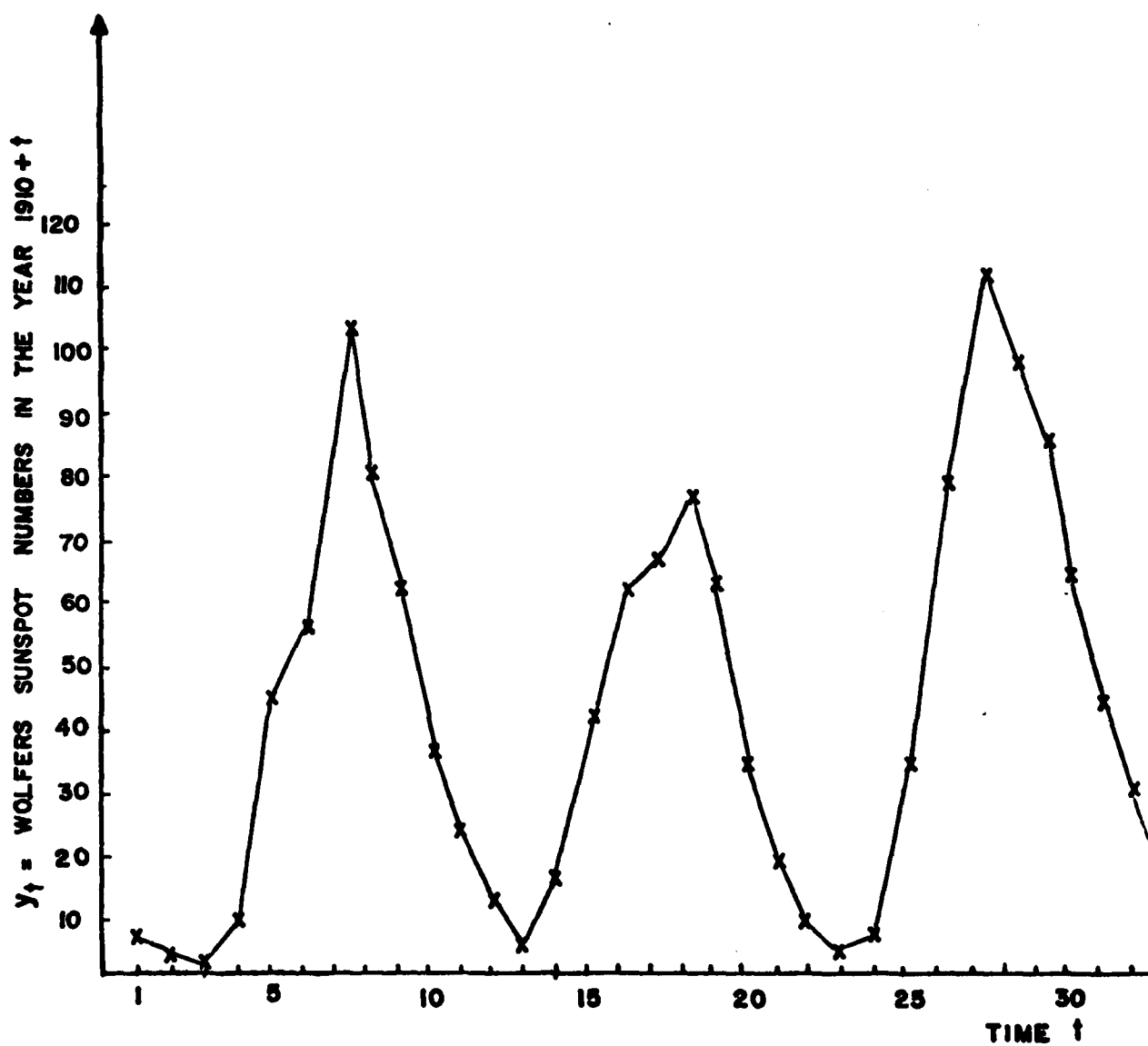


Figure 4.2.1. Wolfer's sunspot numbers against time.

The values of y_t computed by using the above formula are given in column 3 of Table 4.1. We note that these values are quite close to the observed values of y_t ; the small differences between the two sets of values is due to the rounding errors in obtaining the x_i 's, $i = 1, 2, \dots, 33$. These computations were made on a computer using double precision. In practice with a longer series one would use the Fast Fourier Transform. Computer packages are available.

It is of interest to note that the largest pair of coefficients (except for the constant x_1) is $x_6 = 129.425$ and $x_7 = 102.646$. These contribute

$$129.425 \cos 2\pi \frac{t}{11} + 102.646 \sin 2\pi \frac{t}{11} = 162.073 \cos (2\pi \frac{t}{11} - .6705)$$

to the sum. This corresponds to a period of 11 years.

Example 2

In Table 4.2 we present some data on the average bi-monthly expenses^{1/} y_t (in local currency), of a typical family in Kabiria (a city in Northern Algeria) over the time period Jan.-Feb. 1975 through Nov.-Dec. 1977. We would like to obtain a Fourier representation of the y_t , $t = 1, 2, \dots, 18$.

Even though the bi-monthly expenses should constitute a periodic series with a period of 6, we note that due to the randomness of the data $y_t \neq y_{t+6}$, $t = 1, 2, \dots, 12$. Thus we will have to obtain 18 Fourier coefficients x_1, \dots, x_{18} to obtain the desired Fourier representation. (We could of course treat this as periodic data with a period 6 and obtain the 6 Fourier coefficients x_1^*, \dots, x_6^*).

^{1/} Average of the actual expenses for two months.

We shall first obtain

$$x_1 = \frac{1}{\sqrt{18}} \sum_{t=1}^{18} y_t = 22.3$$

$$x_{2k} = \sqrt{\frac{2}{18}} \sum_{t=1}^{18} y_t \cos \frac{2\pi k}{T} t, \quad k = 1, 2, \dots, 8$$

$$x_{2k+1} = \sqrt{\frac{2}{18}} \sum_{t=1}^{18} y_t \sin \frac{2\pi k}{T} t, \quad k = 1, 2, \dots, 8,$$

$$x_{18} = \sqrt{\frac{1}{18}} \sum_{t=1}^{18} y_t (-1)^t = \frac{1}{\sqrt{18}} (10.15) = 2.3923779.$$

The computed values of x_1, \dots, x_{18} are shown in column 4 of Table 4.2. Since $T (= 18)$ is even, the Fourier representation of $y_t, t = 1, 2, \dots, 18$ is given as

$$y_t = \sqrt{\frac{2}{18}} \left[\frac{x_1}{\sqrt{2}} + \sum_{k=1}^8 (x_{2k} \cos \frac{2\pi k}{18} t + x_{2k+1} \sin \frac{2\pi k}{18} t) + x_{18} \frac{(-1)^t}{\sqrt{2}} \right]$$

The values of y_t computed by using the above formula are given in column 3 of Table 4.2. The slight disparity between the observed values y_t (column 2 of Table 4.2), and Fourier representation of y_t (column 3 of Table 4.2), is due to the rounding and computational errors in obtaining the $x_i, i = 1, 2, \dots, 18$.

In Figure 4.3, we show a plot of the observed values of y_t .

Table 4.2

The Average Bi-Monthly Expenses of a Family in
Kabiria and Their Fourier Representation

Time t	Average Bi-Monthly Expenses, y_t	Fourier Represen- tation of y_t	Fourier Coefficients x_k Corresponding to Frequency $k/18$	
			x_k	k
1 Jan-Feb 75	4.71	4.71586	22.304505	1
2 Mar-Apr 75	3.80	3.79525	.254477	2
3 May-Jun 75	3.33	3.33307	.269736	3
4 Jul-Aug 75	9.50	9.49889	.285116	4
5 Sept-Oct 75	6.21	6.20894	-.777919	5
6 Nov-Dec 75	4.27	4.27305	-1.22643	6
7 Jan-Feb 76	4.34	4.3353	-6.17766	7
8 Mar-Apr 76	4.31	4.31574	.364677	8
9 May-Jun 76	3.65	3.64386	-.066686	9
10 Jul-Aug 76	9.67	9.67573	-.903802	10
11 Sep-Oct 76	5.33	5.32531	-.151617	11
12 Nov-Dec 76	3.00	3.00306	-4.96687	12
13 Jan-Feb 77	5.31	5.3089	4.34145	13
14 Mar-Apr 77	3.34	3.33897	.709414	14
15 May-Jun 77	3.36	3.363	.652794	15
16 Jul-Aug 77	10.5	10.4953	.00203447	16
17 Sept-Oct 77	6.00	6.00566	-.414406	17
18 Nov-Dec 77	4.00	3.99402	2.3923779	18

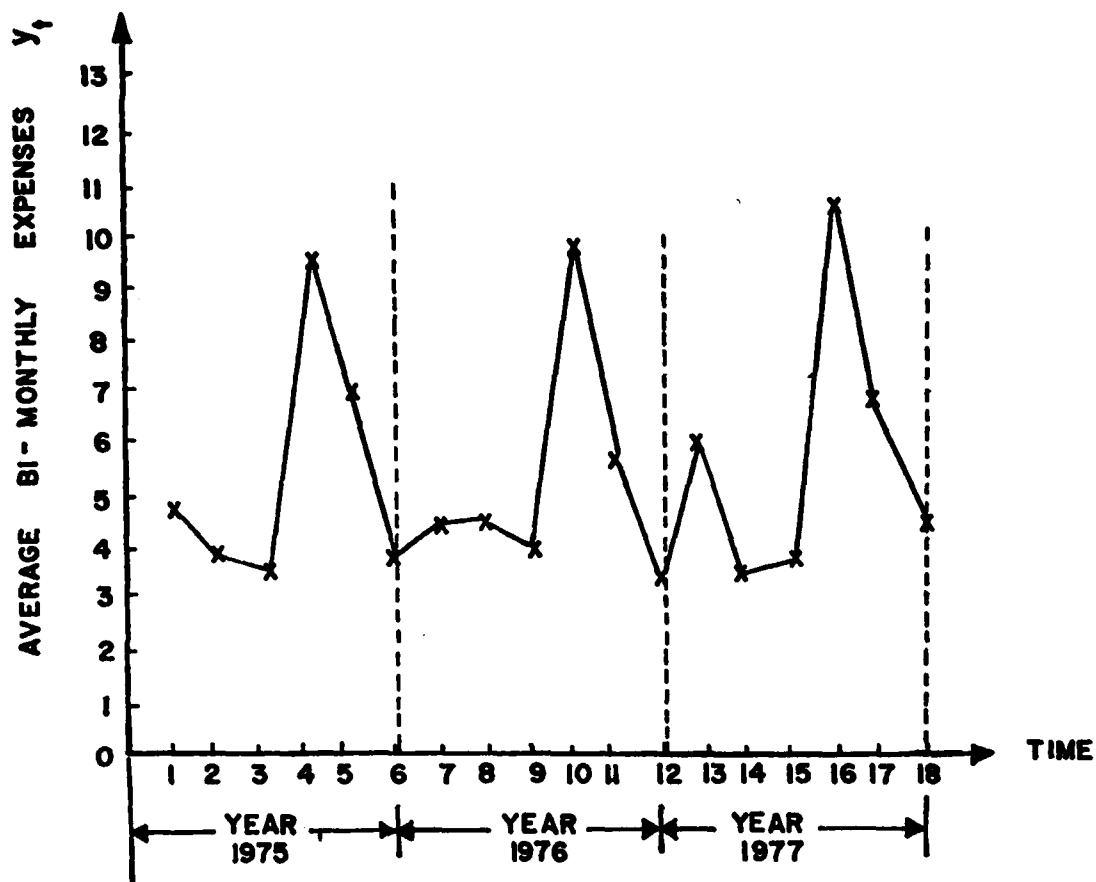


Figure 4.3. Average bi-monthly expenses of a family in Kabiria against time.

Here the two largest pairs of coefficients (except the constant x_1) are:

$$x_6 = 1.2264 \quad \text{and} \quad x_7 = -6.1776 \quad , \quad \text{and}$$

$$x_{12} = -4.96687 \quad \text{and} \quad x_{13} = 4.34145 \quad .$$

They contribute

$$-1.2264 \cos \frac{2\pi t}{6} - 6.1776 \sin \frac{2\pi t}{6} = 6.2982 \cos \left(\frac{2\pi t}{6} - 1.3748 \right) \quad ,$$

$$-4.96687 \cos \frac{2\pi t}{3} + 4.34145 \sin \frac{2\pi t}{3} = 6.5968 \cos \left(\frac{2\pi t}{6} + .7183 \right)$$

to the sum. This corresponds to a period of 12 months and 6 months respectively (note, the data is bi-monthly).

We shall reconsider this data, and perform some additional analysis on it in Sections 4.3 and 4.4.

Example 3

Let us consider the periodic function with period 5

$$f(t) = 1 - t^2 \quad , \quad 0 \leq t < 1 \quad ,$$

$$= -1 \quad , \quad 1 \leq t < 3 \quad ,$$

$$= \frac{1}{4}(t - 3)^2 \quad , \quad 3 \leq t < 5 \quad .$$

We would like to approximate this function by a finite sum of Fourier terms with coefficients α_k and β_k for

i) $k = 0, 1$, and

ii) $k = 0, 1, 2, 3$.

Note that the periodic function $f(t)$ is defined for all values of t , $0 \leq t < \infty$. Since the function is periodic with period $\phi = 5$, it suffices for us to concentrate on the range $0 \leq t < 5$. From (4.7), (4.8) and (4.9), we have

$$\begin{aligned}\alpha_0 &= \frac{1}{5} \int_0^5 f(t) dt = \frac{1}{5} \left[\int_0^1 (1-t^2) dt + \int_1^3 (-1) dt + \int_3^5 \frac{1}{4} (t-3)^2 dt \right] \\ &= -\frac{2}{15}\end{aligned}$$

$$\alpha_1 = \frac{2}{5} \int_0^5 f(t) \cos \frac{2\pi}{5} t dt = .9107 ,$$

$$\alpha_2 = \frac{2}{5} \int_0^5 f(t) \cos \frac{2\pi}{5} 2t dt = 0.15 ,$$

$$\alpha_3 = \frac{2}{5} \int_0^5 f(t) \cos \frac{2\pi}{5} 3t dt = 0.10 ,$$

$$\beta_1 = \frac{2}{5} \int_0^5 f(t) \sin \frac{2\pi}{5} t dt = -.3768 ,$$

$$\beta_2 = \frac{2}{5} \int_0^5 f(t) \sin \frac{2\pi}{5} 2t dt = .2033 ,$$

$$\beta_3 = \frac{2}{5} \int_0^5 f(t) \sin \frac{2\pi}{5} 3t dt = .180 .$$

Thus our approximation for $f(t)$ is given by

$$i) f(t) = \alpha_0 + \alpha_1 \cos \frac{2\pi}{5}t + \beta_1 \sin \frac{2\pi}{5}t$$

$$= -.13 + .9107 \cos \frac{2\pi}{5}t - .37680 \sin \frac{2\pi}{5}t, \quad 0 \leq t < 5;$$

$$ii) f(t) = \alpha_0 + \alpha_1 \cos \frac{2\pi}{5}t + \beta_1 \sin \frac{2\pi}{5}t + \alpha_2 \cos \frac{2\pi}{5}2t + \beta_2 \sin \frac{2\pi}{5}2t$$

$$+ \alpha_3 \cos \frac{2\pi}{5}3t + \beta_3 \sin \frac{2\pi}{5}3t$$

$$= -.13 + .9107 \cos \frac{2\pi}{5}t - .37680 \sin \frac{2\pi}{5}t + .15 \cos \frac{2\pi}{5}2t$$

$$+ .2033 \sin \frac{2\pi}{5}2t + .10 \cos \frac{2\pi}{5}3t + .180 \sin \frac{2\pi}{5}3t,$$

$$0 \leq t < 5.$$

In Figure 4.4 we show a plot of $f(t)$ by the boldface line and a plot of the above two approximations to $f(t)$ by the dotted lines. Based upon an examination of these plots one can see that there has been a significant improvement in the approximation in going from $k = 0, 1$, to $k = 0, 1, 2, 3$, the latter approximation reveals a flatter curve at the bottom than the former. In principle, as we increase k the approximation gets better. The computation of the α_k 's and the β_k 's is rather cumbersome and can be best accomplished on a computer.

4.2 Statistical Estimation Procedures for Cyclical Trends

We shall now turn to the main theme of Section 4; that is, the statistical estimation of cyclical trends. Recall that our basic model

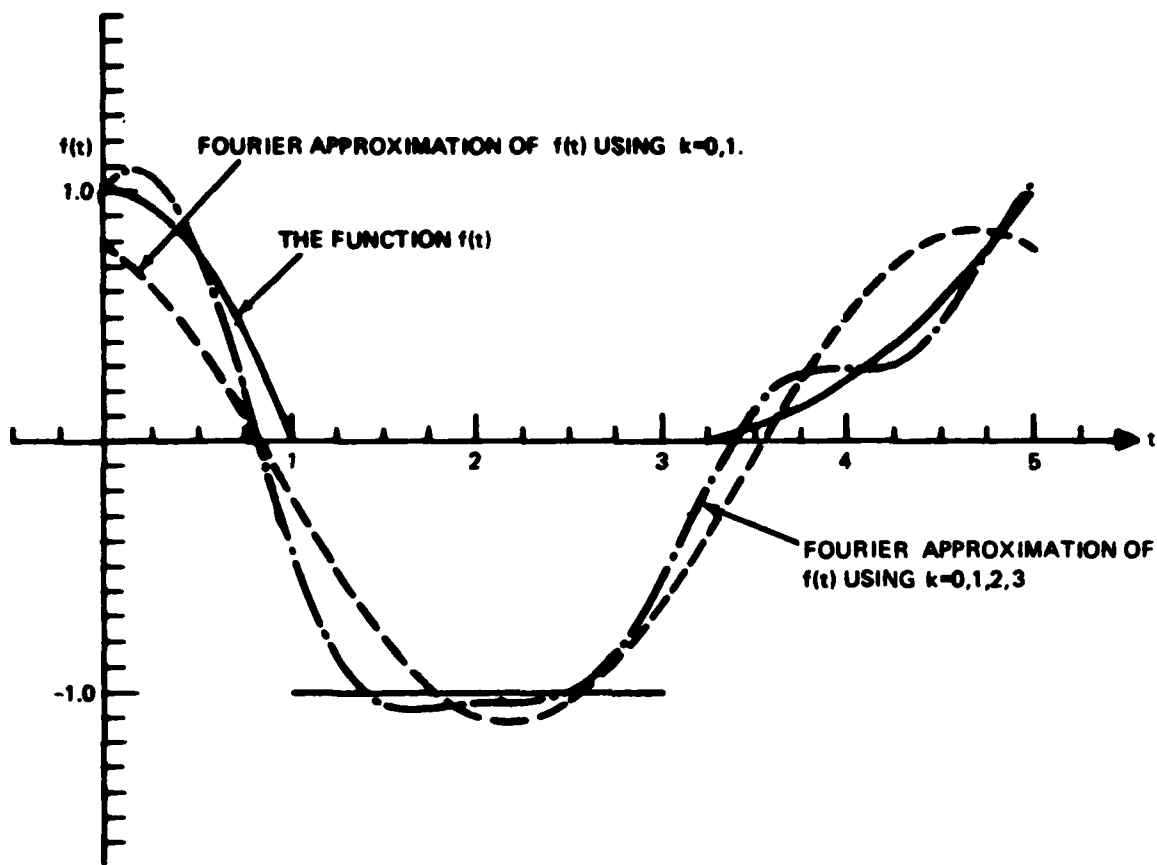


FIGURE 4.4 A PLOT OF THE FUNCTION $f(t)$ AND ITS APPROXIMATIONS USING FOURIER TERMS (WITH $k=0,1$, AND $k=0,1,2,3$).

for an observed time series y_t , $t = 1, 2, \dots, T$, is

$$y_t = f(t) + u_t$$

with $E u_t = 0$, $E u_t^2 = \sigma^2$, and $E u_s u_t = 0$, $t \neq s$. Assume that $f(t)$ is periodic with a known period which divides T . Our goal is to obtain an estimate of $f(t)$ and make some inferences about it using the T observed values y_1, y_2, \dots, y_T .

Since $f(t)$ is periodic, we can represent it by a linear combination of sines and cosines using the methodology of Section 4.1. To this effect, suppose that T is odd, and consider the set of integers $I = \{1, 2, \dots, \frac{T-1}{2}\}$; let $\{k_1, k_2, \dots, k_q\}$ be any proper subset of I . (A proper subset is strictly smaller than the full set.) For example, let T be 9, so that the set of integers I is $\{1, 2, 3, 4\}$; then we can choose $\{k_1 = 1, k_2 = 3, \text{ and } k_3 = 4\}$ as a subset of I . In order to be able to estimate σ^2 (and also test hypotheses about some parameters to be introduced later), it is important that $\{k_1, \dots, k_q\}$ be a proper subset of I , and not be equal to I .

Having chosen the subset k_1, k_2, \dots, k_q , let us consider functions of the form $\sin(\frac{2\pi}{T} k_j t)$ and $\cos(\frac{2\pi}{T} k_j t)$, $j = 1, 2, \dots, q$; the periods of these functions are T/k_j , $j = 1, 2, \dots, q$. Following the development in Section 4.1.3, let us consider $f(t)$ as

$$(4.10) \quad f(t) = \alpha_0 + \sum_{j=1}^q (\alpha(k_j) \cos \frac{2\pi}{T} k_j t + \beta(k_j) \sin \frac{2\pi}{T} k_j t) ,$$

where α_0 , $\alpha(k_j)$ and $\beta(k_j)$ are the coefficients associated with the trigonometric terms. Note that in the above representation, the trigonometric terms with period 2 are not included. In order to include trigonometric terms with a period 2, T must be even; this case will be considered later. If we let

$$\rho(k_j) = \sqrt{\alpha^2(k_j) + \beta^2(k_j)} \quad , \quad \theta(k_j) = \arctan \frac{\beta(k_j)}{\alpha(k_j)} \quad ,$$

then

$$\alpha(k_j) = \rho(k_j) \cos \theta(k_j) \quad \text{and} \quad \beta(k_j) = \rho(k_j) \sin \theta(k_j) \quad ,$$

and so

$$(4.11) \quad f(t) = \alpha_0 + \sum_{j=1}^q \rho(k_j) \cos \left[\frac{2\pi k_j}{T} t - \theta(k_j) \right] \quad .$$

Using the observed values of the series y_1, y_2, \dots, y_T , our objective is to obtain the least squares estimators of α_0 , $\alpha(k_j)$ and $\beta(k_j)$, $j = 1, 2, \dots, q$. We shall write

$$\begin{aligned} y_t = & \alpha_0 + \rho(k_1) \cos \left[\frac{2\pi k_1}{T} t - \theta(k_1) \right] + \rho(k_2) \cos \left[\frac{2\pi k_2}{T} t - \theta(k_2) \right] \\ & + \dots + \rho(k_q) \cos \left[\frac{2\pi k_q}{T} t - \theta(k_q) \right] + u_t \quad , \quad t = 1, 2, \dots, T \quad , \end{aligned}$$

as our model.

By taking advantage of the orthogonality property of the trigonometric terms, and by using standard techniques, the least squares estimators of α_0 , $\alpha(k_j)$ and $\beta(k_j)$ are:

$$(4.12) \quad \alpha_0 = \frac{1}{T} \sum_{t=1}^T y_t = \bar{y}$$

$$(4.13) \quad \alpha(k_j) = \frac{2}{T} \sum_{t=1}^T y_t \cos \frac{2\pi k_j}{T} t, \quad j = 1, 2, \dots, q,$$

$$(4.14) \quad \beta(k_j) = \frac{2}{T} \sum_{t=1}^T y_t \sin \frac{2\pi k_j}{T} t, \quad j = 1, 2, \dots, q.$$

Because of orthogonality, the above estimators are uncorrelated; also the least squares estimator of σ^2 takes a particularly simple form

$$(4.15) \quad s_1^2 = \frac{\sum_{t=1}^T y_t^2 - T\bar{y}^2 - \frac{1}{2}T \sum_{j=1}^q [\alpha^2(k_j) + \beta^2(k_j)]}{T - (2q + 1)},$$

and the estimates of $\rho(k_j)$ and $\theta(k_j)$ are

$$(4.16) \quad R(k_j) = \sqrt{\alpha^2(k_j) + \beta^2(k_j)},$$

$$(4.17) \quad \hat{\theta}(k_j) = \arctan \frac{\beta(k_j)}{\alpha(k_j)}, \quad j = 1, 2, \dots, q$$

respectively.

When T is even, we consider subsets of the integers $\{1, 2, \dots, \frac{T}{2}\}$. For $T/2$ we have $\cos \frac{2\pi T/2}{T} t = \cos \pi t = (-1)^t$ and $\sin \frac{2\pi T/2}{T} t = \sin \pi t = 0$; there is only one function. When T is even, we can include the coefficient associated with a period 2, writing $f(t)$ by

$$f(t) = \alpha_0 + \sum_{j=1}^q [\alpha(k_j) \cos \frac{2\pi k_j}{T} t + \beta(k_j) \sin \frac{2\pi k_j}{T} t] + \alpha_{T/2} (-1)^t .$$

The least squares estimator of $\alpha_{T/2} (-1)^t$ is

$$(4.18) \quad a_{T/2} = \frac{1}{T} \sum_{t=1}^T (-1)^t y_t .$$

Thus, the estimator of $f(t)$ is

$$(4.19) \quad \hat{f}(t) = a_0 + \sum_{j=1}^q [a(k_j) \cos \frac{2\pi}{T} k_j t + b(k_j) \sin \frac{2\pi}{T} k_j t] ,$$

if T is odd, and

$$\hat{f}(t) = a_0 + \sum_{j=1}^q [a(k_j) \cos \frac{2\pi}{T} k_j t + b(k_j) \sin \frac{2\pi}{T} k_j t] + a_{T/2} (-1)^t ,$$

if T is even.

An estimate of σ^2 when T is even is given by

$$(4.20) \quad s_2^2 = \frac{\sum_{t=1}^T y_t^2 - T(\bar{y})^2 + a_{T/2}^2 - \frac{1}{2}T \sum_{j=1}^q [a^2(k_j) + b^2(k_j)]}{T - (2q + 2)} .$$

Remarks:

We note that the expressions for a_0 , $a(k_j)$, and $b(k_j)$ given above are similar to those of the discrete Fourier coefficients x_1, x_2, \dots, x_T of a sequence of numbers y_1, \dots, y_T . Specifically, $a(k) = \sqrt{T/2} x_{2k}$ and $b(k) = \sqrt{T/2} x_{2k+1}$. In fact, if we had to choose $\{k_1, k_2, \dots, k_q\} = \{1, 2, \dots, (T-1)/2\}$ (when T is odd), then the estimated values of a_0 , $a(k_j)$ and $b(k_j)$ would be proportional to the discrete Fourier coefficients, so that $y_t = \hat{f}(t)$. However, this choice of the $\{k_j\}$ will not permit us to estimate σ^2 and test hypotheses about the $\alpha(k_j)$ and the $\beta(k_j)$.

4.3 The Periodogram and the Spectrum

There are two distinct ideas which motivate our study of cyclical trends. One is to describe seasonal variation where the periods are known. The other is to discover "hidden periodicities". The periodogram and the spectrum are important tools which enable us to identify the periodic components in a time series and provide us with a vehicle to interpret visually the estimates $a(k_j)$ and $b(k_j)$.

Note that $\hat{f}(t)$ given by equation (4.19) can also be written as

$$(4.21) \quad \hat{f}(t) = a_0 + \sum_{j=1}^q R(k_j) \cos \left(\frac{2\pi}{T} k_j t - \hat{\theta}(k_j) \right), \quad T \text{ odd}$$

$$= a_0 + \sum_{j=1}^q R(k_j) \cos \left(\frac{2\pi}{T} k_j t - \hat{\theta}(k_j) \right) + a_{T/2} (-1)^t, \quad T \text{ even}.$$

The quantity $R(k_j)$ given in (4.16) denotes the estimated amplitude of the cosine curve with frequency k_j/T and period T/k_j .

A plot of $R^2(k_j)$ versus T/k_j (the period) is called the periodgram. When T is odd, the periodgram may be defined for the periods $T, T/2, T/3, \dots, 2T(T-1)$, whereas if T is even, the periodgram may be defined for the periods $T, T/2, T/3, \dots, 2T$. In either case, the periodgram is defined for periods greater than or equal to 2.

A plot of $R^2(k_j)$ versus k_j/T (the frequency) is called the spectrogram.^{1/} It is defined for the frequencies $\frac{1}{T}, \frac{2}{T}, \dots, \frac{1}{2}$ if T is even, and for the frequencies $\frac{1}{T}, \frac{2}{T}, \dots, \frac{1}{2} - \frac{1}{2T}$ if T is odd. In either case, the spectrogram is defined for frequencies less than or equal to a half. The frequencies being more evenly spaced than the periods makes the spectrogram more convenient to use than the periodgram.

4.3.1 Interpretation of the Spectrum

Since $a(k_j)$ and $b(k_j)$ are the least squares estimators, they are the values of $\alpha(k_j)$ and $\beta(k_j)$ that minimize (for odd T)

$$\sum_{t=1}^T [y_t - (\alpha(k_j) \cos \frac{2\pi k_j}{T} t + \beta(k_j) \sin \frac{2\pi k_j}{T} t)]^2,$$

and thus $R(k_j) = \sqrt{a^2(k_j) + b^2(k_j)}$ and $\hat{\theta}(k_j) = \arctan \frac{b(k_j)}{a(k_j)}$ are the values of $\rho(k_j)$ and $\theta(k_j)$ that minimize

^{1/} In the literature there is confusion about these terms. Many authors use "periodgram" to mean the graph of amplitude against frequency.

$$\sum_{t=1}^T [y_t - \rho(k_j) \cos(\frac{2\pi}{T} k_j t - \theta(k_j))]^2 .$$

The minimum value of this sum of squares is

$$\sum_{t=1}^T y_t^2 - \frac{1}{2} T (a^2(k_j) + b^2(k_j)) = \sum_{t=1}^T y_t^2 - \frac{1}{2} T R^2(k_j) .$$

In view of the above, we conclude that $R(k_j)$ is a measure of how closely a trigonometric function with frequency k_j/T fits the observed data. More pragmatically, if a series of length T has period ϕ , then the value of $R(k_j)$ corresponding to $k_j = T/\phi$ will tend to be largest among all the other $R(k_j)$'s.

In practice, a plot of the spectrogram enables us to detect the periods in the time series by our identifying the frequencies k_j/T associated with large values of $R(k_j)$.

In this section we consider the spectrogram as giving information about the trend. In a later section we shall interpret the spectrogram in terms of the Fourier transform of the correlation function.

4.3.2 Example Illustrating the Estimation of a Cyclical Trend and the Spectrogram

We shall illustrate the methodology of Sections 4.2 and 4.3 by considering the data in Table 4.2 on the average bi-monthly expenses of a family in Kabiria. The sample Fourier coefficients and the estimated amplitudes are given in Table 4.2.1. (Note $a(h) = \sqrt{T/2} x_{2h} = 3x_{2h}$, $b(h) = \sqrt{T/2} x_{2h+1} = 3x_{2h+1}$, $h = 1, \dots, 8$, and $a_9 = \sqrt{18} x_{18}$.)

The estimated amplitudes are graphed in Figure 4.4.1.

It seems reasonable to consider seasonal variation; that is, mean value function can be suspected as having a period of 6 since

Table 4.2.1

The Discrete Fourier Coefficients for the Data in Table 4.2 on the Average Bi-Monthly Expenses

Values of k_j	$a(k_j)$	$b(k_j)$	$R^2(k_j)$	$R(k_j)$
1	.0848	.0894	.0152	.1236
2	.0950	-.2593	.0762	.2761
3	-.4088	-2.0592	4.4075	2.0994
4	.1215	-.0222	.0152	.1235
5	-.3012	-.0505	.0933	.3054
6	-1.6556	1.4471	4.8353	2.1989
7	.2364	.2176	.1032	.3213
8	.0067	-.1381	.0193	.1383
9	.5639		.3180	

$$a_0 = \bar{y} = 5.2572, \quad \sum_{t=1}^{18} (y_t - \bar{y})^2 = 10.2012$$

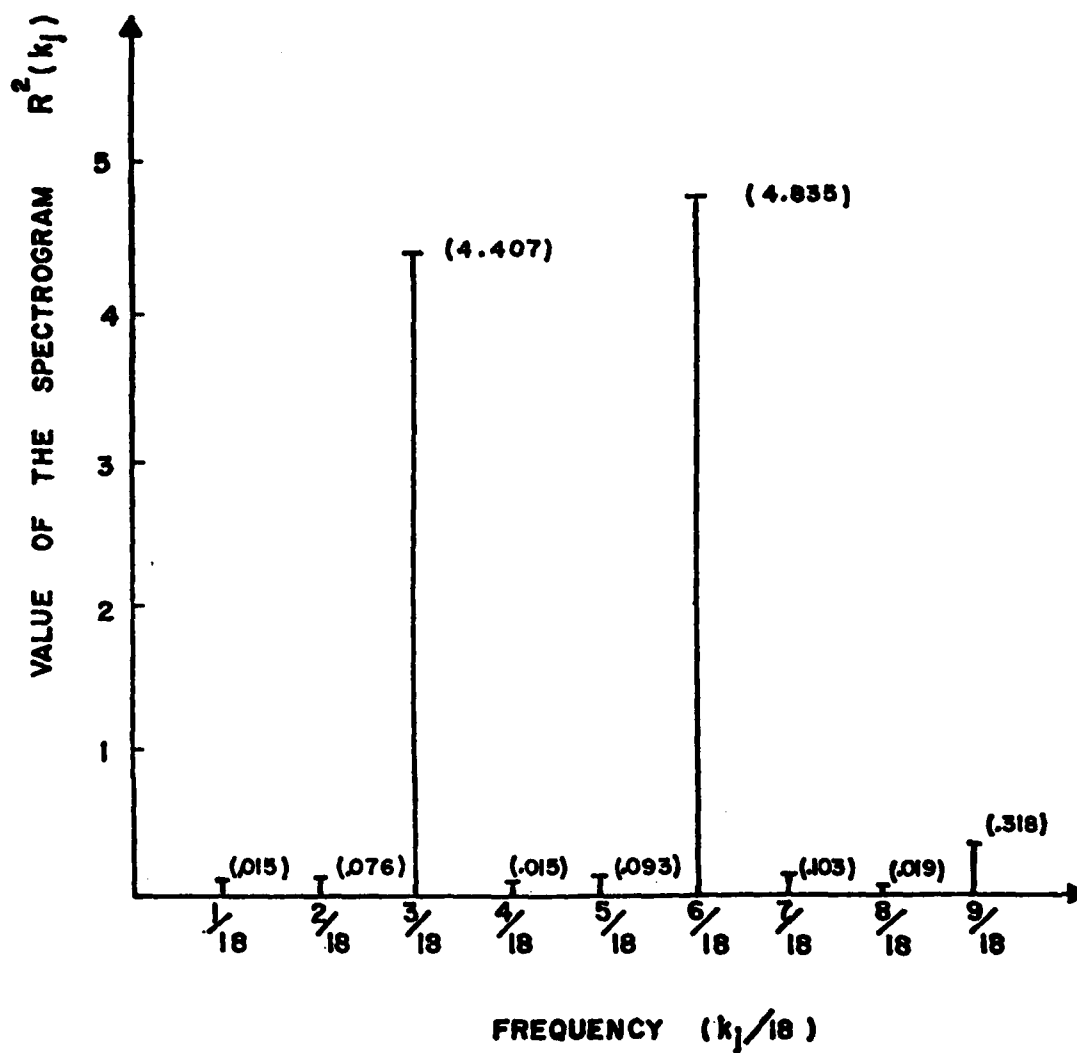


Figure 4.4.1. A spectrogram of the average bi-monthly expenses of a family in Kabiria.

the data are bi-monthly. The trigonometric functions with period 6 correspond to the integers 3, 6, and 9; that is frequencies $1/6$, $1/3$, and $1/2$. We consider

$$f(t) = \alpha_0 + \sum_{j=1}^2 [\alpha(3j) \cos \frac{2\pi}{18} 3jt + \beta(3j) \sin \frac{2\pi}{18} 3jt] + \alpha_9(-1)^t .$$

It will be seen in Figure 4.4.1 that the estimated amplitudes at frequencies $1/6$, $2/6$, and $3/6$ are considerably larger than at the other frequencies. This fact confirms that the seasonal variation is dominant. We estimate the variance by

$$s^2 = \frac{\sum_{t=1}^{18} y_t^2 - 18a_0^2 - 9 \sum_{j=1}^2 R^2(3j) - 18a_9^2}{12}$$

$$= .2418 ,$$

and $s = .4917$.

4.4 Tests of Hypotheses and Confidence Regions for Coefficients

Because they are the least squares estimates, $\mathbb{E}(a_0) = \alpha_0$, $\mathbb{E}(a_{T/2}) = \alpha_{T/2}$, $\mathbb{E}(a(k_j)) = \alpha(k_j)$, and $\mathbb{E}(b(k_j)) = \beta(k_j)$, $j = 1, 2, \dots, q$. The variances of a_0 and $a_{T/2}$ are σ^2/T , whereas the variances of $a(k_j)$ and $b(k_j)$ are $2\sigma^2/T$. To verify the latter, let us recall that

$$a(k_j) = \frac{2}{T} \sum_{t=1}^T y_t \cos \frac{2\pi k_j}{T} t ,$$

and from the orthogonality of the cosine function (see equation (4.1)), the variance of $a(k_j)$ is $2\sigma^2/T$; similarly, the variance of $b(k_j)$ is $2\sigma^2/T$.

If we assume that the y_t 's are normally distributed, then, since the estimators a_0 , $a_{T/2}$, $a(k_j)$ and $b(k_j)$ are uncorrelated, they are also normally and independently distributed. Let $s_1^2 = s^2$ for T odd and $s_2^2 = s^2$ for T even. Then $\frac{(T - 2q - 1)s_1^2}{\sigma^2}$ has

a chi-square distribution with $(T - 2q - 1)$ degrees of freedom, and

$\frac{(T - 2q - 2)s_2^2}{\sigma^2}$ has a chi-square distribution with $(T - 2q - 2)$ degrees of freedom.

A null hypothesis which may be of interest is whether there exists a cyclical term with a minimum period T/k_j for some specific j . Thus, we wish to test the null hypothesis H_0 , that

$$\alpha(k_j) = \beta(k_j) = 0 , \text{ or equivalently}$$

$$\rho(k_j) = 0$$

versus the alternate hypothesis that $\rho(k_j) \neq 0$.

Under the null hypothesis both the $a(k_j)$ and the $b(k_j)$ are independent and are distributed normally with mean 0 and variance $2\sigma^2/T$. Thus $a^2(k_j)/(2\sigma^2/T)$ and $b^2(k_j)/(2\sigma^2/T)$ are independent and each has

AD-A092 575

STANFORD UNIV CA DEPT OF STATISTICS F/G 12/1
METHODS AND APPLICATIONS OF TIME SERIES ANALYSIS. PART I. REGRE--ETC(U)
JUL 80 T W ANDERSON; N D SINGPURWALLA N00014-75-C-0442

UNCLASSIFIED

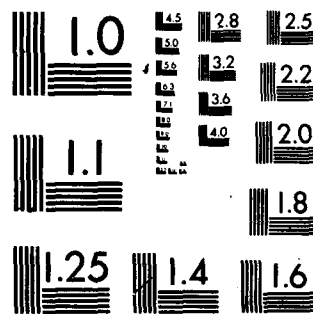
TR-42

NL

2
100



END
DATE
FILMED
1-8
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

a chi-square distribution with one degree of freedom. It follows then, that $(a^2(k_j) + b^2(k_j))T/(2\sigma^2)$ has a chi-square distribution with 2 degrees of freedom; equivalently, $R^2(k_j)T/2\sigma^2$ has a chi-square distribution with 2 degrees of freedom.

To complete our discussion on testing the hypothesis H_0 , let us denote by p , the number of coefficients that we have estimated. Specifically,

$$\begin{aligned} p &= 2q + 1, \text{ if } T \text{ is odd, and} \\ &= 2q + 2, \text{ if } T \text{ is even and the term } \alpha_{T/2} \text{ is estimated by us.} \end{aligned}$$

Since $\frac{(T-p)s_i^2}{\sigma^2}$ has a chi-square distribution with $(T-p)$ degrees of freedom, the ratio

$$\frac{\frac{T}{2} \frac{R^2(k_j)}{\sigma^2} / 2}{\frac{(T-p)s_i^2}{\sigma^2} / (T-p)} = \frac{T}{4s_i^2} R^2(k_j), \quad i = 1, 2,$$

has an F-distribution with 2 and $(T-p)$ degrees of freedom. We can use this result to test the null hypothesis H_0 using standard procedures.

If we want to test the hypothesis $\rho(k_j) = 0, j = 1, \dots, q$, $\alpha_{T/2} = 0$, when T is even, we use

$$\frac{\frac{T}{2} \sum_{j=1}^q R^2(k_j) + T \alpha_{T/2}^2}{(2q+1)s_2^2},$$

which has the F-distribution with $2q + 1$ and $T - 2q - 2$ degrees of freedom, respectively, when the null hypothesis is true.

4.4.1 Example

Let us consider again the bi-monthly expenditures. The hypothesis that there is no seasonal variation is the hypothesis that

$$\alpha_3 = \beta_3 = \alpha_6 = \beta_6 = \alpha_9 = 0 .$$

The appropriate test criterion is

$$\frac{9[R^2(3) + R^2(6)] + 18a_9^2}{5s^2} = \frac{91.81}{5 \times .2418} = 75.94 .$$

This is to be referred to the F-distribution with 5 and 12 degrees of freedom. It is clearly significant. (This is obvious from Figure 4.4.1.)

Given that there is seasonal variation, we can ask whether all of the 5 terms are needed. The criterion for testing $\rho(3) = 0$ is

$$\frac{18 \times 4.4075}{4 \times .3224} = 61.52 ,$$

which is obviously significant. To test the null hypothesis that

$\alpha_9 = 0$, we use

$$\frac{\sqrt{T} a_9}{s} = \frac{\sqrt{18} \times .5639}{.4917} = 4.8656 ,$$

which is referred to the t-distribution with 12 degrees of freedom.

It is significant. ($t_{12}(.01) = 3.054.$)

We could ask whether any of the coefficients with periods 1/18, 2/18, 4/18, 5/18, 7/18, 8/18 are needed. However, the corresponding estimated amplitudes are so small compared to the seasonal amplitudes that they cannot be important (even if statistically significant).

The above considerations lead us to propose $\hat{f}(t)$, where $f(t)$ is a trigonometric function of period 6, as our estimate of the cyclical trend $f(t)$. Specifically,

$$\begin{aligned}\hat{f}(t) = & 5.257 - .409 \cos \frac{2\pi}{18} 3t - 2.059 \sin \frac{2\pi}{18} 3t - 1.656 \cos \frac{2\pi}{18} 6t \\ & + 1.447 \sin \frac{2\pi}{18} 6t + .564(-1)^t, \quad t = 1, \dots, 18.\end{aligned}$$

In Table 4.2.2 we give the computed values $\hat{f}(t)$ together with the actual values y_t , and the residuals $y_t - \hat{f}(t)$, $t = 1, \dots, 18$.

In Figure 4.4 we give a plot of the actual and the fitted values y_t and $\hat{f}(t)$, respectively. This enables us to judge visually the appropriateness of $\hat{f}(t)$. Figure 4.5 is a plot of the residuals $(y_t - \hat{f}(t))$ against time t . The analysis of residuals is an important part of a statistical study, since the residuals give us an indication of how well the fitted values $\hat{f}(t)$ behave in relationship to the actual values y_t . In Figure 4.6, we plot the fitted values $\hat{f}(t)$ against the observed values y_t ; such a plot is also a part of the statistical analysis.

Table 4.2.2

Actual Bi-Monthly Expenses of a Family in Kabiria and
Their Fitted Values Using a Trigonometric Function of Period 6

Time t	Actual Average Bi-Monthly Expenses y_t	Fitted Values $\hat{f}(t)$	Residuals $y_t - \hat{f}(t)$
1	4.71	4.787	-.077
2	3.80	3.817	-.017
3	3.33	3.447	-.117
4	9.50	9.889	-.400
5	6.21	5.847	.370
6	4.27	3.757	.513
7	4.34	4.787	-.447
8	4.31	3.817	.493
9	3.65	3.446	.204
10	9.67	9.889	-.220
11	5.33	5.847	-.517
12	3.00	3.757	-.757
13	5.31	4.787	.523
14	3.34	3.817	-.477
15	3.36	3.447	-.087
16	10.50	9.889	.610
17	6.00	5.847	.153
18	4.00	3.757	.243

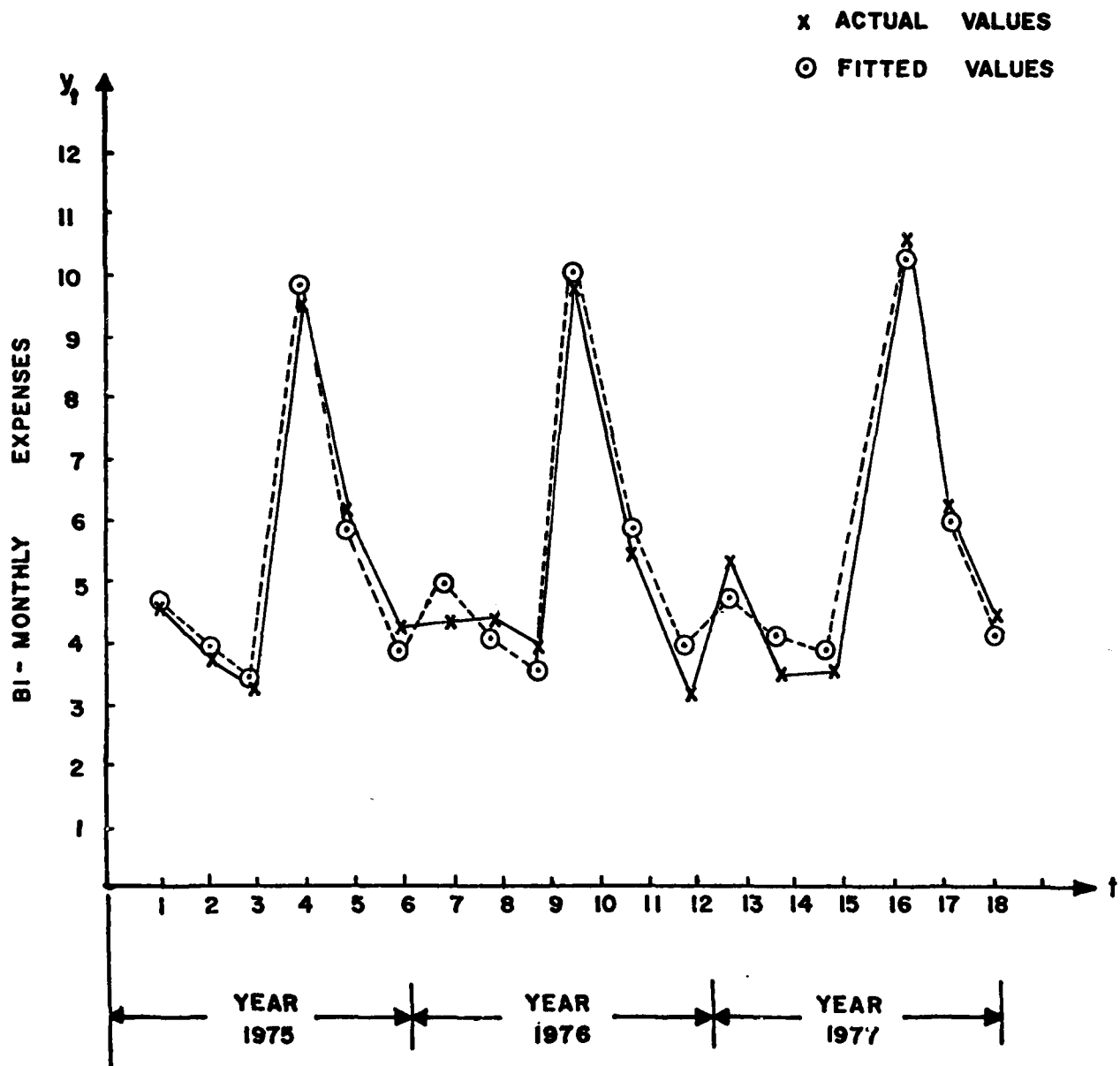


Figure 4.4. Average bi-monthly expenses of a family in Kabiria and their fitted values using a trigonometric function of period 6.

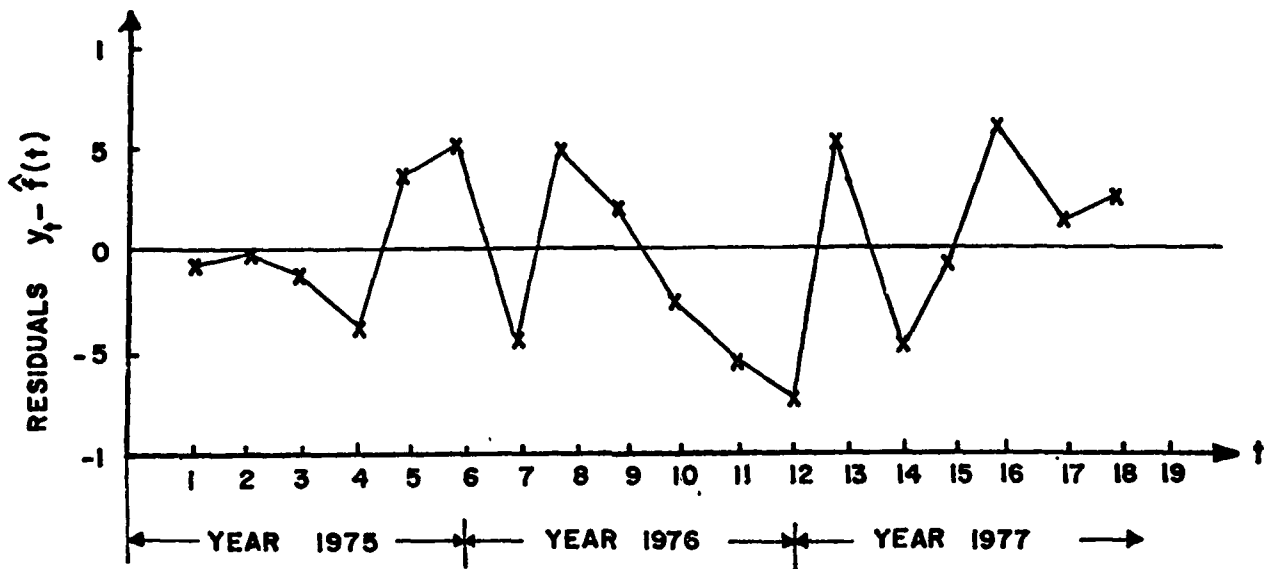


Figure 4.5. Residuals ($y_t - \hat{f}(t)$) against time for average bi-monthly expenses data.

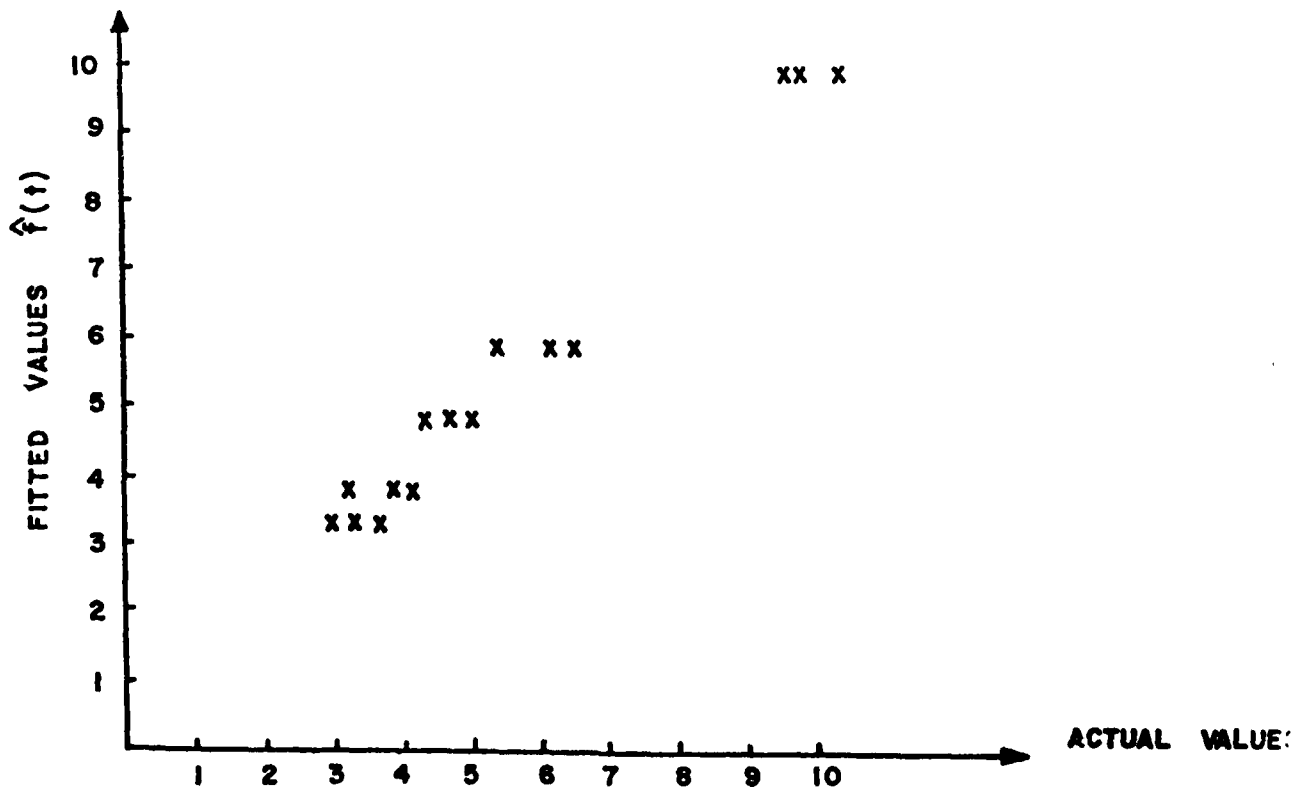


Figure 4.6. Predicted (fitted) values against actual values for average bi-monthly expenses data.

References

- Anderson, T.W. (1971), "The Statistical Analysis of Time Series," John Wiley and Sons, Inc., New York.
- Beals, Ralph E. (1972), Statistics for Economists: An Introduction, Rand McNally Company, Chicago.
- Bhattacharyya, G.K. and Klotz, J.H. (1966), "The Bivariate Trend of Lake Mendota," Technical Report No. 98, Department of Statistics, University of Wisconsin, Madison, Wisconsin.
- Biometrika Tables for Statisticians (1962), Vol. I (Edited by E.S. Pearson and H.O. Hartley), Cambridge University Press.
- Fisher, R.A. and Yates, Frank (1963), "Statistical Tables for Biological, Agricultural and Medical Research (Sixth Edition), Oliver and Boyd Ltd., Edinburgh.
- Mann, N.R., Schafer, R.E., and Singpurwalla, N.D. (1974), "Methods for Statistical Analysis of Reliability and Life Data," John Wiley and Sons, Inc., New York.
- Singpurwalla, N.D. (1975), "Times Series Analysis and Forecasting of Failure Rate Processes," Reliability and Fault Free Analysis (edited by Barlow, R.E., et al) SIAM, Philadelphia.

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.
2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.
3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.
4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.
5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.
6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.
7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.
8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.
9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.
10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.
11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.
12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.
13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.
14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.
15. "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.
16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time series," Steffen L. Lauritzen, April, 1974.

TECHNICAL REPORTS (continued)

17. "General Exponential Models for Discrete Observations,"
Steffen L. Lauritzen, May, 1974.
18. "On the Interrelationships among Sufficiency, Total Sufficiency and
Some Related Concepts," Steffen L. Lauritzen, June, 1974.
19. "Statistical Inference for Multiply Truncated Power Series Distributions,"
T. Cacoullos, September 30, 1974.

Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)

20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models
in the Time and Frequency Domains," T. W. Anderson, June 1975.
21. "Asymptotic Properties of Some Estimators in Moving Average Models,"
Raul Pedro Mentz, September 8, 1975.
22. "On a Spectral Estimate Obtained by an Autoregressive Model Fitting,"
Mituaki Huzii, February 1976.
23. "Estimating Means when Some Observations are Classified by Linear
Discriminant Function," Chien-Pai Han, April 1976.
24. "Panels and Time Series Analysis: Markov Chains and Autoregressive
Processes," T. W. Anderson, July 1976.
25. "Repeated Measurements on Autoregressive Processes," T. W. Anderson,
September 1976.
26. "The Recurrence Classification of Risk and Storage Processes,"
J. Michael Harrison and Sidney I. Resnick, September 1976.
27. "The Generalized Variance of a Stationary Autoregressive Process,"
T. W. Anderson and Raul P. Mentz, October 1976.
28. "Estimation of the Parameters of Finite Location and Scale Mixtures,"
Javad Behboodian, October 1976.
29. "Identification of Parameters by the Distribution of a Maximum
Random Variable," T. W. Anderson and S.G. Ghurye, November 1976.
30. "Discrimination Between Stationary Gaussian Processes, Large Sample
Results," Will Gersch, January 1977.
31. "Principal Components in the Nonnormal Case: The Test for Sphericity,"
Christine M. Waternaux, October 1977.
32. "Nonnegative Definiteness of the Estimated Dispersion Matrix in a
Multivariate Linear Model," F. Pukelsheim and George P.H. Styan, May 1978.

TECHNICAL REPORTS (continued)

33. "Canonical Correlations with Respect to a Complex Structure," Steen A. Andersson, July 1978.
34. "An Extremal Problem for Positive Definite Matrices," T.W. Anderson and I. Olkin, July 1978.
35. "Maximum likelihood Estimation for Vector Autoregressive Moving Average Models," T. W. Anderson, July 1978.
36. "Maximum likelihood Estimation of the Covariances of the Vector Moving Average Models in the Time and Frequency Domains," F. Ahrabi, August 1978.
37. "Efficient Estimation of a Model with an Autoregressive Signal with White Noise," Y. Hosoya, March 1979.
38. "Maximum Likelihood Estimation of the Parameters of a Multivariate Normal Distribution," T.W. Anderson and I. Olkin, July 1979.
39. "Maximum Likelihood Estimation of the Autoregressive Coefficients and Moving Average Covariances of Vector Autoregressive Moving Average Models," Fereydoon Ahrabi, August 1979.
40. "Smoothness Priors and the Distributed Lag Estimator," Hirotugu Akaike, August, 1979.
41. "Approximating Conditional Moments of the Multivariate Normal Distribution," Joseph G. Deken, December 1979.
42. "Methods and Applications of Time Series Analysis - Part I: Regression, Trends, Smoothing, and Differencing," T.W. Anderson and N.D. Singpurwalla, July 1980.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 14 TR-42	2. GOVT ACCESSION NO. AD-A092575	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) METHODS AND APPLICATIONS OF TIME SERIES ANALYSIS. PART I. REGRESSION, TRENDS, SMOOTHING, AND DIFFERENCING.	5. TYPE OF REPORT & PERIOD COVERED Technical Report	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) A. W. ANDERSON and N. D. SINGPURWALLA	8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0442	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-034)	
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Code 436 Arlington, Virginia 22217	12. REPORT DATE JUL 1980	13. NUMBER OF PAGES 98
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED	16. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) time series analysis, trends, smoothing, cyclical trends, regression		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This is the first in a series of technical reports developing the most modern procedures of time series analysis and forecasting for use in engineering, the physical sciences, and the social sciences. The exposition of methodology is based on a succinct presentation of the theoretical background and is illustrated with appropriate examples from engineering, maintenance and reliability, economics, and other physical and social sciences. The first is concerned with Regression, Trends, Smoothing, and Differencing.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

332580

JW